ALGEBRAIC TRACE FUNCTIONS OVER THE PRIMES

ÉTIENNE FOUVRY, EMMANUEL KOWALSKI, AND PHILIPPE MICHEL

Abstract. We study sums over primes of trace functions of ℓ-adic sheaves. Using an extension of our earlier results on algebraic twist of modular forms to the case of Eisenstein series and bounds for Type II sums based on similar applications of the Riemann Hypothesis over finite fields, we prove general estimates with power-saving for such sums. We then derive various concrete applications.

Contents

1. Introduction

Let $f(X) = P(X)/Q(X)$ $P, Q \in \mathbf{Z}(X)$ be a non-zero rational function. If p is a prime large enough so that $f(X)$ defines a rational function on $\mathbf{F}_p$ by reduction modulo p, it follows from the work of Weil that we have the estimate

$$\sum_{\substack{1\leqslant n\leqslant p\\ (Q(n),p)=1}} e\Big(\frac{P(n)\overline{Q(n)}}{p}\Big) \ll p^{1/2}$$

(where $Q(n)\overline{Q(n)} = 1 \,(\mathrm{mod}\, p)$) which exhibits considerable cancellation in this exponential sum. It is a natural question, with many potential applications, to ask whether such cancellation persists when the sum is restricted to prime numbers q, either less than p or over a shorter intervals (longer intervals being usually easier to handle).

In [14], Fouvry and Michel proved that this is indeed almost always the case:

Theorem 1.1. *Assume that $f = P/Q$ is not a polynomial of degree $\leqslant 1$. Then we have*

$$\sum_{\substack{q<p \ prime\\ (Q(q),p)=1}} e\Big(\frac{P(q)\overline{Q(q)}}{p}\Big) \ll X\Big(\frac{p}{X}\Big)^{7/32} p^{-\eta}$$

Date: November 27, 2012, 2:09.

2010 *Mathematics Subject Classification.* 11N05, 11N13, 11N32, 11N35, 11F11, 11T23, 11L05.

Key words and phrases. Sums over primes, Möbius function, Eisenstein series, trace functions of ℓ-adic sheaves, Riemann Hypothesis over finite fields.

Ph. M. was partially supported by the SNF (grant 200021-137488) and the ERC (Advanced Research Grant 228304). É. F. thanks ETH Zürich, EPF Lausanne and the Institut Universitaire de France for financial support.

for $X \leqslant p$ and for any $\eta < 1/32$, where the implicit constant depends only on the degrees of P and Q and on η.

Similar estimates were already known when f is a polynomial (of degree > 1), but with the exponent η depending on the degree of f and tending to 0 as the latter increased (see, e.g, [21, 24]). Thus an important new feature in [14] was to allow for the most general possible rational fractions, and for a uniform p-power saving. One of the key input of the proof was an essential use of Deligne's theory of higher dimensional algebraic exponential sums.

There are, however, many other functions defined over $\mathbf{F}_p$ for which one would like to have similar results. For instance, with $f(X) = P(X)/Q(X)$, one may naturally want to consider

$$K(n) = \begin{cases} \chi(f(n)) & \text{if } (p, Q(n)) = 1\ , \\ 0 & \text{if } p|Q(n), \end{cases}$$

for some non-trivial Dirichlet character χ modulo p of order $h \geqslant 2$, provided the rational function f is not an h-th power. Nevertheless, the only example we are aware of concerns the case where $f(X) = aX+b$, with $b \neq 0$, is a polynomial of degree 1, which was considered first by Karatsuba [28] (and later [18]) who obtained a positive power saving, which is non trivial whenever $X \geqslant p^{1/2+\varepsilon}$ for some $\varepsilon > 0$. In Corollary 1.11, we will prove a non-trivial bound for an (almost) arbitrary rational function f.

Beyond additive and multiplicative characters, there are other function defined over finite fields which are now common tools in number theory. A nice example is given by the (normalized) hyper-Kloosterman sums, introduced by Deligne and studied by Katz in great detail in [29], which are defined by

$$K(n) = \mathrm{Kl}_m(n;p) = \frac{1}{p^{\frac{m-1}{2}}} \sum \cdots \sum_{\substack{x_1 \cdots x_m = n \\ x_i \in \mathbf{F}_p}} e\Big(\frac{x_1 + \cdots + x_m}{p}\Big),$$

for some integer $m \geqslant 2$. This example was first considered by the third author in [32–34], who obtained a modest (yet non-trivial) saving of $\frac{\log \log p}{\log p}$ over the trivial bound $O(p/\log p)$ for the sum over primes $q < p$ of $\mathrm{Kl}_m(q;p)$.

Remark 1.2. One can also wonder about the very recent generalizations of Kloosterman sums $\mathrm{Kl}^{\varrho}_{\check{\mathbf{G}}}$ associated to the general Kloosterman sheaves defined, for an arbitrary split reductive group $\check{\mathbf{G}}$ and a representation ϱ of it, by Heinloth, Ngô and Yun [23], where the case of the hyper-Kloosterman sums above corresponds to $\check{\mathbf{G}} = \mathrm{GL}_n$ with its standard representation. It is a sign of the generality of our results that they do apply very straightforwardly to this case, although the corresponding trace functions have not (yet) been explicited for all[1] $\check{\mathbf{G}}$!

The common link between all these functions is that they are special cases of the general class of functions we called *trace weights* in [12], with bounded conductors. Precisely, we have the following definition:

Definition 1.3 (Trace weights)**.** For a prime p and a prime $\ell \neq p$, an *isotypic trace sheaf modulo* p is a geometrically isotypic ℓ-adic Fourier sheaf $\mathcal{F}$ on $\mathbf{A}^1_{\mathbf{F}_p}$, in the sense of [29, Def. 8.2.2], which is pointwise pure of weight 0.

An *isotypic trace weight modulo* p is the trace function

$$K(x) = \iota((\mathrm{tr}\,\mathcal{F})(\mathbf{F}_p, x))$$

[1] As Ngô kindly informed us, Yun has computed these sums explicitly for $\check{\mathbf{G}} = \mathrm{SO}(2n+1)$ and its standard representation: $\mathrm{Kl}_{\mathrm{SO}(3)}$ is the symmetric square of Kl_2 and $\mathrm{Kl}_{\mathrm{SO}(2n+1)}$ for $n \geqslant 2$ is essentially the multiplicative convolution of $\mathrm{Kl}_{\mathrm{SO}(3)}$ and of two Kloosterman sums Kl_n, see [40].

for $x \in \mathbf{F}_p$ of an isotypic trace sheaf $\mathcal{F}$, this trace function being seen as complex-valued by means of some fixed isomorphism $\iota : \bar{\mathbf{Q}}_\ell \to \mathbf{C}$.

To any constructible sheaf $\mathcal{F}$ on $\mathbf{A}^1_{\bar{\mathbf{F}}_p}$ is associated its *analytic conductor*, a numerical invariant which measures the complexity of $\mathcal{F}$. This is a positive integer defined by

$$\mathrm{cond}(\mathcal{F}) = \mathrm{rank}(\mathcal{F}) + \sum_x (1 + \mathrm{Swan}_x(\mathcal{F})),$$

where x ranges over the (finitely many) singularities of $\mathcal{F}$ in $\mathbf{P}^1(\overline{\mathbf{F}_p})$, i.e., those x where $\mathcal{F}$ is not lisse, and $\mathrm{Swan}_x(\mathcal{F}) \geqslant 0$ is the Swan conductor of $\mathcal{F}$ at x (see [29]). For an isotypic trace weight $K(n)$, we define the conductor as the minimal conductor of an isotypic trace sheaf $\mathcal{F}$ with trace function equal to $K(n)$ on $\mathbf{F}_p$.

Remark 1.4. For example:

- If $K(n) = e(P(n)/p)$ for a polynomial P of degree $< p$, the associated sheaf has conductor $\mathrm{cond}(\mathcal{F}) = \deg P + 2$;
- If $K(n) = \chi(P(n))$ where χ is multiplicative and P a polynomial, then it is bounded by 2 plus the number of distinct zeros of P in $\bar{\mathbf{F}}_p$;
- For the hyper Kloosterman sums $K(n) = \mathrm{Kl}_m(n; p)$, the conductor is $m + 3$;
- For the trace function of the ℓ-adic Kloosterman sheaf associated to the adjoint representation of the split reductive group $\check{\mathbf{G}}$ in [23], the conductor is bounded by $\dim(\check{\mathbf{G}}) + 2 + r(\check{\mathbf{G}})$, where $r(\check{\mathbf{G}})$ is the rank of $\check{\mathbf{G}}$ (by [23, p. 4, (3)]: this sheaf is of dimension $\dim \mathrm{Ad} = \dim \check{\mathbf{G}}$, lisse on $\mathbf{G}_m$, tame at 0 and with Swan conductor $r(\check{\mathbf{G}})$ at ∞).

Our main result in this paper is an estimate for any sum over primes q of an isotypic trace function modulo p, which is universal in quality and gives power-saving whenever the length of the sum is roughly comparable with p on a logarithmic scale, in particular allowing some sums over shorter intervals $q < p^{1-\theta}$ for some $\theta > 0$. The only weights we can not handle are those where the corresponding estimate would be tantamount to a “quasi-Riemann Hypothesis”, i.e., a zero-free strip for some Dirichlet L-functions.

It will therefore be natural to say that $K(n)$ is an *exceptional weight modulo p* (for sums over primes) if it is proportional to a weight

$$K_{\chi,\psi}(n) = \chi(n)\psi(n)$$

where χ (resp. ψ) is a multiplicative (resp. additive) character modulo p, where either or both may be trivial.

Similarly, a sheaf $\mathcal{F}$ will be called *exceptional* if it is geometrically isotypic and geometrically isomorphic to a sum of copies of a tensor product $\mathcal{L}_\chi \otimes \mathcal{L}_\psi$ of a Kummer sheaf with an Artin-Schreier sheaf, so that its trace function is exceptional.

We state the results both for standard and for smoothed sums over primes. We will consider smooth test functions V, compactly supported in $[1/2, 2]$, such that

$$x^j V^{(j)}(x) \ll Q^j \tag{1.1}$$

for some $Q \geqslant 1$ and for any integer $j \geqslant 0$, where the implicit constant depends on j.

Theorem 1.5 (Trace weights vs. primes)**.** *Let K be an isotypic trace weight on $\mathbf{F}_p$ associated to some sheaf $\mathcal{F}$, and assume that $\mathcal{F}$ is not exceptional. Let V be a smooth function as above satisfying* (1.1) *for some parameter $Q \geqslant 1$.*

For any $X \geqslant 2$, we have

$$\sum_{q \text{ prime}} K(q) V\Big(\frac{q}{X}\Big) \ll QX(1+p/X)^{1/6} p^{-\eta}, \tag{1.2}$$

$$\sum_{\substack{q \text{ prime} \\ q \leqslant X}} K(q) \ll X(1+p/X)^{1/12} p^{-\eta/2}, \tag{1.3}$$

for any $\eta < 1/24$. The implicit constants depend only on η, $\mathrm{cond}(\mathcal{F})$ *and the implicit constants in* (1.1)*. Moreover, the dependency on* $\mathrm{cond}(\mathcal{F})$ *is at most polynomial.*

Remark 1.6. For $X = p$ one gets

$$\sum_{\substack{q \text{ prime} \\ q < p}} K(q) \ll p^{1-1/48+\varepsilon},$$

and for general X these bounds are non-trivial as long as the conductor of $\mathcal{F}$ remains bounded and the range X is greater that $p^{3/4+\varepsilon}$ for some $\varepsilon > 0$. Stronger results are available by different methods for special K. For instance, Bourgain [3] and Bourgain-Garaev [5] have obtained bounds for

$$K(n) = e\Big(\frac{an + b\overline{n}^k}{p}\Big),\ k \in \mathbf{N} - \{0\},\ (b,p) = 1,$$

which are non-trivial as long as $X \geqslant p^{1/2+\varepsilon}$ (see also [17] for a survey of existing methods).

Closely related to Theorem 1.5 is the following estimate:

Theorem 1.7 (Trace weights vs. Möbius)**.** *Let μ denote the Möbius function. With the same notations and hypotheses as in Theorem 1.5, we have for $X \geqslant 2$*

$$\sum_{n} \mu(n) K(n) V\Big(\frac{n}{X}\Big) \ll QX(1+p/X)^{1/6} p^{-\eta},$$

$$\sum_{n \leqslant X} \mu(n) K(n) \ll X(1+p/X)^{1/12} p^{-\eta/2},$$

for any $\eta < 1/24$, where the implicit constants depend only on η, $\mathrm{cond}(\mathcal{F})$ *and the implicit constants in* (1.1)*, and the dependency on* $\mathrm{cond}(\mathcal{F})$ *is at most polynomial.*

Remark 1.8. J. Bourgain pointed out that Theorem 2 of [6] (along with the Note following it) can be used in conjonction with (3.2) and Proposition 3.1 of the present paper to prove that if K is an isotypic trace weight, we have, for any $\varepsilon > 0$ and for any $X \geqslant p^{1/2+\varepsilon}$

$$\sum_{n \leqslant X} \mu(n) K(n) = o(X),$$

where the implicit constant depends on $\mathrm{cond}(\mathcal{F})$ and ε. However, this approach does not seem to yield a power saving, and does not seem to apply if μ is replaced by Λ or by the characteristic function of the primes.

Remark 1.9. This theorem expresses an orthogonality property (i.e., the absence of correlation) between the Möbius function and any isotypic trace weight modulo p with bounded conductor. This fits with the philosophy of the "Möbius randomness principle", formulated vaguely in [26, p. 338], and with Sarnak's recent precise formulation in terms of orthogonality of the Möbius function against function with low complexity, in the sense of entropy (see [37]). Our result is in a slightly different context than Sarnak's conjecture, however, since the trace weights $K(n)$ are defined modulo p, and an asymptotic statement follows only by taking, for each p, a different weight with some bound on the complexity, as measured by the conductor in our case. Thus, our results are closer in spirit to those of Green [20] and Bourgain [4], which prove asymptotic orthogonality of the Möbius function against, respectively, bounded depth boolean functions, and monotone boolean functions on binary hypercubes $\{0,1\}^N$ identified with $\{1,\ldots,2^N\}$.

In fact, it seems to be a very intriguing question (suggested by Sarnak) to understand which functions can arise as small linear combinations of trace functions of low conductor (which, by linearity, still satisfy Theorems 1.5 and 1.7). Another natural question is whether trace functions have low complexity in a more algorithmic sense, and this does not seem to be easy to answer. Only functions such as

$$K(n)=e\Bigl(\frac{P(n)}{p}\Bigr) \text{ or } K(n)=\Bigl(\frac{P(n)}{p}\Bigr)$$

(for $P(X)\in\mathbf{Z}[X]$ fixed and $(\frac{\cdot}{p})$ the Legendre symbol) seem to be obviously of low complexity for most meaning of the term, and such functions as

$$K(n)=\mathrm{Kl}_2(n;p)$$

are far from being understood in this respect. One can certainly expect the class of trace weights (and linear combinations with small coefficients) to be very rich and fascinating (in this respect, there are already hints in Deligne's conjecture about the number of trace functions with various conditions, see [8, 9, 13] for this topic.)

1.1. **First applications.** We present here some corollaries of Theorem 1.5 which are obtained by applying it to specific weights K. This is only a selection and we expect many more applications. We leave to the reader to write down the corresponding statements involving the Möbius function, which follow from Theorem 1.7.

In the first result, we obtain a power saving in the sum of the error term in the prime number theorem in arithmetic progressions over residues classes modulo a prime which are values of some fixed polynomial. Precisely, we define $E(X;p,a)$ by

$$\pi(X;p,a)=\frac{\delta_p(a)}{\varphi(p)}\pi(X)+E(X;p,a),$$

where $\delta_p(a)=0$ if $(a,p)\neq 1$ and is 1 otherwise.

We will prove in §5.1 :

Corollary 1.10. *Let $P\in\mathbf{Z}[X]$ be a polynomial whose reduction modulo p is squarefree and non-constant.*

(1) *We have*

$$\sum_{n\in\mathbf{F}_p}E(X;p,P(n))\ll X(1+p/X)^{1/12}p^{-\eta}$$

for any $\eta<1/48$, where the implicit constant depends only on η and $\deg P$*.*

(2) *We have*

$$\sum_{a\in P(\mathbf{F}_p)}E(X;p,a)\ll X(1+p/X)^{1/12}p^{-\eta}$$

for any $\eta<1/48$, where the implicit constant depends only on η and $\deg P$*.*

The restriction to a squarefree polynomial could be relaxed, but some condition is needed in the current state of knowledge since for $P=X^2$ we have the interpretation

$$\sum_{n\in\mathbf{F}_p}E(X;p,n^2)=\sum_{q\leqslant X}\Bigl(\frac{q}{p}\Bigr)+O\Bigl(\frac{X}{p}+1\Bigr)$$

in terms of average of the Legendre symbol over primes, from which we can not get power saving without using a quasi-Riemann Hypothesis for the corresponding L-function.

On the other hand, if we take $P = X^2 - 1$, the study of either of these sums becomes equivalent to that of the sum

$$\sum_{q<X} \chi(q+1)$$

where q, again, runs over primes. This was estimated, as we recalled, by Karatsuba [28].

We can now generalize considerably this result of Karatsuba:

Corollary 1.11 (Character sums over polynomially-shifted primes)**.** *Let $f = P/Q$ be a rational function represented as a ratio of integral polynomials. Let χ be a non-trivial Dirichlet character of prime modulus p and order $h \geqslant 2$. Assume that f modulo p is not of the form*

$$cX^k g(X)^h$$

for some $c \in \mathbf{F}_p{}^{\times}$, some $k \in \mathbf{Z}$ and some $g(X) \in \mathbf{F}_p(X)$. We then have

$$\sum_{q \ \textit{prime}} \chi(f(q))V(q/X) \ll X(1+p/X)^{1/6}p^{-\eta}$$

$$\sum_{\substack{q \ \textit{prime} \\ q \leqslant X}} \chi(f(q)) \ll X(1+p/X)^{1/12}p^{-\eta/2}$$

for any $\eta < 1/24$, where the implicit constant depends only on η, V and the degrees of P and Q.

Proof. The weight $K(n) = \chi(f(n))$ is associated with the tame, rank-1, Kummer sheaf $\mathcal{L}_{\chi(f)}$ which has (at most) $\deg P + \deg Q$ singularities, hence conductor bounded in terms of $\deg P$ and $\deg Q$. It can only be exceptional if it is geometrically isomorphic to $\mathcal{L}_{\chi^k} \simeq \mathcal{L}_{\chi(X^k)}$ for some $k \geqslant 0$. But this is well-known to be precisely equivalent with $f(X) = cX^k g(X)^h$. □

We leave to the reader the statement of a result unifying this corollary with Theorem 1.1 for

$$K(n) = \chi(f(n))e\Big(\frac{g(n)}{p}\Big).$$

1.2. **Kloosterman sums at prime arguments.** Our last application involves the weights $K(n)$ which are related to Kloosterman or hyper-Kloosterman sums Kl_m. We first spell out two very specific corollaries for the standard Kloosterman sum in one variable:

Corollary 1.12. *For every $0 < \eta < 1/48$ there exists $C(\eta)$ such that for every p, every $X \geqslant 2$ and every integer n coprime with p, one has the inequalities*

$$\Big| \sum_{q<X,\, q \ \textit{prime}} \mathrm{Kl}_2(nq;p) \log q \Big| \leqslant C(\eta) X(1+p/X)^{1/12}p^{-\eta}$$

and

$$\Big| \sum_{q<X,\, q \ \textit{prime}} \mathrm{Kl}_2(n^2q^2;p) e\Big(\frac{2nq}{p}\Big) \log q \Big| \leqslant C(\eta) X(1+p/X)^{1/12}p^{-\eta}.$$

These two bounds improve [27, Lemmas 6.1, 6.2, 6.3] when $c = p$ is a prime and when X is near and possibly a bit smaller than p. Using the methods of [27], one can use the second bound in conjunction and the Petersson formula to increase the size of the support of the Fourier transform $\widehat{\Phi}$ of the test functions Φ in the problem of computing the distribution of low-lying zeros (average 1-level density) of the symmetric square L-functions $L(\mathrm{sym}^2 f, s)$ for f in the family of holomorphic newforms of prime level $p \to +\infty$ and weight k: with notation as in [27] (except that they denote the level by N), *there exists $\delta > 0$ such that for any $\Phi \in \mathcal{S}(\mathbf{R})$ with the support of $\widehat{\Phi}$ in $]-1/2-\delta, 1/2+\delta[$, one has*

$$\lim_{p\to\infty} \frac{1}{|H_k^\star(p)|} \sum_{f\in H_k^\star(p)} D(\mathrm{sym}^2 f, \Phi) = \int_{\mathbf{R}} \Phi(x) W(\mathrm{Sp})(x) dx.$$

The possibility of such an improvement was known to the authors of [27] (see [27, Remark C, p. 61]), though their method was different.

The consideration of powers of hyper-Kloosterman sums allows us to strengthen the results [33, 34] concerning the existence of hyper-Kloosterman sums with large absolute value modulo a product of two primes:

Corollary 1.13. *For any $m \geqslant 2$, there exists a constant $\alpha_m > 0$ such that*

$$\sum_{c\leqslant X} \Lambda_2(c) |\operatorname{Kl}_m(1;c)| \geqslant (\alpha_m + o_m(1)) X \log X,$$

were $\Lambda_2(c) = (\mu \star \log^2)(c)$ denotes the von Mangoldt function of order 2, which is supported over integers with at most two prime factors.

This corollary shows that the normalized hyper-Kloosterman sums $\operatorname{Kl}_m(1;c)$ whose modulus is a product of at most two primes have their size $\gg_m 1$ for a positive proportion of such moduli (when these are weighted by Λ_2). In [33, 34], the lower-bound was of order X, and by adding the missing logarithmic factor, we obtain the right order of magnitude. This answers a question of Bombieri to the third author from 1996. For Λ_2 replaced by Λ_3, a corresponding (easier) statement was proven in [16].

Another potential application of this corollary (or rather of the techniques used to prove it) is to reduce the value of the constant ω in the following statement, which was first established by Fouvry and Michel for $\omega = 23$ in [15] subsequently improved to $\omega = 18$ by Sivak-Fischler [38, 39] and to $\omega = 15$ by Matomäki [31]:

Theorem. *The sequence $(\operatorname{Kl}_2(1;c))_{c\geqslant 1}$ changes sign infinitely often as c varies over squarefree moduli with at most ω prime factors.*

1.3. Principle of the proof: the combinatorics of sums over primes.

We start from a general perspective before explaining what features are specific to our case and what our new ingredients are. Thus we assume given a bounded arithmetic function K, a smooth function V, compactly supported in $]0, +\infty[$ and some $X \geqslant 2$, and we wish to obtain non-trivial bounds for the sum

$$\sum_n \Lambda(n) K(n) V\Big(\frac{n}{X}\Big),$$

where Λ denotes the von Mangoldt function.

Using Heath-Brown's identity (see, e.g., [26, Prop. 13.3]) and a smooth partition of unity, this sum decomposes essentially into a linear combination of sums of the shape

$$\sum\sum_{m_1,\cdots,m_k} \alpha_1(m_1)\cdots\alpha_k(m_k) \sum\sum_{n_1,\cdots,n_k} V_1(n_1)\cdots V_k(n_k) \\ V\Big(\frac{m_1\cdots m_k n_1\cdots n_k}{X}\Big) K(m_1\cdots m_k n_1 \cdots n_k) \tag{1.4}$$

for some integral parameter $k \geqslant 2$, where the $\alpha_i(m)$ are essentially bounded arithmetic functions supported in dyadic intervals (say $[M_i/2, M_i]$) of short range (i.e. $M_i \leqslant X^{1/k}$), whereas the $V_i(n)$

are smooth functions supported in dyadic intervals with arbitrary range (say, $[N_i/2, N_i]$ with $N_i \in [1/2, 2X]$), and where

$$\prod_i M_i N_i \asymp X.$$

We refer to the n_i as the "smooth" variables and the m_i as the "non-smooth" variables, as one is usually unable to exploit the specific shape of the functions α_i, except for the fact that they are supported in short ranges.

Depending on which estimates and methods are available to bound these sums, according to the location of the point $(M_1, \cdots, M_k, N_1, \cdots, N_k)$ in the $2k$-dimensional cube $[1/2, 2X]^{2k}$, it is useful to classify them into different (not necessarily disjoint) categories, based on the number of "long" smooth variables which are available:

- If there is one very long smooth variable, say n_1, one usually speaks of *sums of type I*, with the remaining (smooth and non-smooth variables) combined together into a single non-smooth variable, m, which means that the original sum (1.4) may be written
$$\sum_{m \asymp M} \beta_m \sum_{n_1 \asymp N_1} V_1(n_1) V\Big(\frac{mn_1}{X}\Big) K(mn_1).$$
- If there are two relatively long smooth variables, say n_1, n_2, one speaks of sums of type I_2; after combining the remaining (smooth and non-smooth variables) into a single non-smooth variable, the sum can now be rewritten
$$\sum_{m \asymp M} \alpha_m \sum\sum_{\substack{n_1 \asymp N_1 \\ n_2 \asymp N_2}} V_1(n_1) V_2(n_2) V\Big(\frac{mn_1n_1}{X}\Big) K(mn_1n_2).$$
- And if there are three relatively long smooth variables, say n_1, n_2, n_3, we will speak of sums of type I_3, and so on.

This classification appears more or less explicitly in the work [10] of Fouvry, in the context of the average distribution of primes in arithmetic progressions to large moduli. The implementation of this strategy depends on the possibility of dealing with the sums of type I_r for r as large as possible, a question which becomes increasingly difficult as r increases, since the range of the smooth variables decreases.[2] All remaining sums belong then to the class of *sums of type II*. The most direct treatment of these sums –there may be other treatments available, depending on the original problem– consists in combining these (short) variables in subsets to form variables with larger ranges, in order to obtain bilinear forms involving two non-smooth variables of the type

$$\sum_{m \asymp M} \sum_{n \asymp N} \alpha_m \beta_n K(mn), \qquad \text{where } MN \asymp X.$$

One can then "smoothen" one of the variables, say n, by an application of the Cauchy-Schwarz inequality, leading to a quadratic form with coefficients with *multiplicative correlation sums* of the function K, namely

$$\sum_{m_1, m_2} \overline{\alpha_{m_1}} \alpha_{m_2} \sum_n W\Big(\frac{n}{N}\Big) \overline{K(m_1 n)} K(m_2 n).$$

Notice here that the fact that the original variables are rather short actually helps, since it offers some flexibility in the ways they may be combined to tailor the relative ranges of M and N. This is the strategy we will follow in this paper.

[2] For instance, in [10], it is shown that one could prove results on the distribution of primes in long arithmetic progressions on average, beyond the Bombieri-Vinogradov Theorem, if one could treat the corresponding sums of type I_r for $r = 1, \ldots, 6$. Currently, the sums of type I_1, I_2 and I_3 can be handled [19].

1.4. **Sums of type I_2.** We can now come to our specific situation and explain our new results for sums over primes of trace weights.

We will give estimates for sums of type I, I_2 and II. In fact, the starting point of this work is a very general estimate for sums of type I_2 (two long smooth variables of approximately equal size) when K is a trace weight, which follows relatively easily from the results of our earlier paper [12]. Indeed, using Mellin inversion, the estimation of sums of type I_2 can be reduced to that of sums of the shape

$$\mathcal{S}_{V,X}(it,K)=\sum_{n}K(n)d_{it}(n)V\Bigl(\frac{n}{X}\Bigr) \tag{1.5}$$

where $t\in\mathbf{R}$, and (for any $u\in\mathbf{C}$) we denote by

$$d_u(n)=d_{-u}(n)=\sum_{ab=n}\Bigl(\frac{a}{b}\Bigr)^{u}$$

the twisted divisor function.

We observe that the arithmetic function $n\to d_{it}(n)$ is (up to suitable normalization) the Fourier coefficient of the non-holomorphic unitary Eisenstein series

$$E(z,s)=\frac{1}{2}\sum_{(c,d)=1}\frac{y^{s}}{|cz+d|^{2s}},$$

for $s=\frac{1}{2}+it$. The main result of our previous paper ([12, Thm 1.2]) is a universal non-trivial bound for the analogue of $\mathcal{S}_{V,X}(it,K)$ where $d_{it}(n)$ is replaced with the Fourier coefficients of a classical cusp form (either holomorphic or not). We will extend the proof to Eisenstein series, obtaining the following result:

Theorem 1.14 (Algebraic twists of Eisenstein series)**.** *Let K be an isotypic trace weight associated to the ℓ-adic sheaf $\mathcal{F}$ modulo p. Let V be a smooth function satisfying* (1.1) *with parameter $Q\geqslant 1$. If $\mathcal{F}$ is not geometrically trivial, then for any $X\geqslant 1$, we have*

$$\mathcal{S}_{V,X}(it,K)=\sum_{n}K(n)d_{it}(n)V\Bigl(\frac{n}{X}\Bigr)\ll(1+|t|)^{A}QX\Bigl(1+\frac{p}{X}\Bigr)^{1/2}p^{-\eta}$$

for any $\eta<1/8$ and some $A\geqslant 1$ possibly depending on η. The implicit constant depends only on η, on the implicit constants in (1.1)*, and polynomially on the conductor of $\mathcal{F}$.*

In fact, the proof of this theorem will be intertwined with the proof of the following estimate on sums of type I_2:

Theorem 1.15 (Type I_2 sums of trace weights)**.** *Let K be an isotypic trace weight associated to the ℓ-adic sheaf $\mathcal{F}$ modulo p. Let $M,N,X\geqslant 1$ be parameters with $X/4\leqslant MN\leqslant X$. Let U, V, W be smooth functions satisfying condition* (1.1) *with respective parameters Q_U,Q_V and Q_W. We then have*

$$\sum_{m,n}K(mn)\Bigl(\frac{m}{n}\Bigr)^{it}U\Bigl(\frac{m}{M}\Bigr)V\Bigl(\frac{n}{N}\Bigr)W\Bigl(\frac{mn}{X}\Bigr)\ll(Q_U+Q_V)^{B}(1+|t|)^{A}Q_WX\Bigl(1+\frac{p}{X}\Bigr)^{1/2}p^{-\eta}$$

for $t\in\mathbf{R}$ and for any $\eta<1/8$ and some constants $A,B\geqslant 1$ depending on η only. The implicit constant depends only on η, on the implicit constants in (1.1)*, and polynomially on the conductor of $\mathcal{F}$.*

Remark. (1) Through the techniques of [12], this result depends on deep results of algebraic geometry, including Deligne's general form of the Riemann Hypothesis over finite fields, and the theory of the ℓ-adic Fourier transform of Deligne, Laumon and Katz.

(2) The Polya-Vinogradov method would yield a non trivial bound for the sum above as long as $\max(M,N) \gg p^{1/2}\log p$. Here we obtain non trivial estimates for $MN \gg p^{3/4+\varepsilon}$ in particular when $M, N \gg p^{3/8+\varepsilon}$.

1.5. **Sums of type I and II.** Our second main result is a general estimate for sums of type II, which gives non-trivial bounds, as long as one of the variables has range slightly greater than $p^{1/2}\log p$ and the other has non-trivial range. Precisely:

Theorem 1.16. *Let K be a* non-exceptional *trace weight modulo p associated to an isotypic ℓ-adic sheaf $\mathcal{F}$. Let $M, N \geqslant 1$ be parameters, and let $(\alpha_m)_m$, $(\beta_n)_n$ be sequences supported on $[M/2, 2M]$ and $[N/2, 2N]$ respectively.*

(1) *We have*

$$\sum\sum_{\substack{m,n\\(m,p)=1}} \alpha_m\beta_n K(mn) \ll \|\alpha\|\|\beta\|(MN)^{1/2}\Bigl(\frac{1}{p^{1/4}} + \frac{1}{M^{1/2}} + \frac{p^{1/4}\log^{1/2}p}{N^{1/2}}\Bigr), \tag{1.6}$$

where

$$\|\alpha\|^2 = \sum_m |\alpha_m|^2,\ \ \|\beta\|^2 = \sum_n |\beta_n|^2.$$

(2) *We have*

$$\sum_{(m,p)=1} \alpha_m \sum_{n\leqslant N} K(mn) \ll \Bigl(\sum_m |\alpha_m|\Bigr) N\Bigl(\frac{1}{p^{1/2}} + \frac{p^{1/2}\log p}{N}\Bigr). \tag{1.7}$$

In both estimates, the implicit constants depend only, and at most polynomially, on the conductor of $\mathcal{F}$.

This theorem constitutes a significant generalization of results like [33, Cor. 2.11] or [14, Prop. 1.3], which were obtained for very specific weights (additive characters of rational functions or symmetric powers of Kloosterman sums). The main difference is that we do not require any knowledge of the geometric monodromy group of $\mathcal{F}$. Instead, it turns out that we can build on the same ideas used in [12] to handle algebraic twists of cusp forms. A crucial role is played again by the ℓ-adic Fourier transform, and by a geometric invariant of $\mathcal{F}$ which we introduced in [12], namely its *Fourier-Möbius group*, which controls the correlation of the trace function of the Fourier transform of $\mathcal{F}$ with its pullbacks under automorphisms of the projective line. In fact, it is the intersection of this group with the standard Borel subgroup of upper-triangular matrices of $\mathrm{PGL}_2(\mathbf{F}_p)$ which we must understand, the essential point being that this intersection is of size bounded in terms of the conductor of $\mathcal{F}$ *unless* $\mathcal{F}$ is exceptional. This is the origin of this restriction in Theorem 1.16. It is rather remarkable that the upper-triangular matrices in the Fourier-Möbius group were precisely those which do *not* cause any difficulty in [12] (hence in Theorem 1.15).

Remark 1.17. For the purpose of Theorem 1.5, it is indeed enough to handle sums of type I_2 and to deal will all others as sums of type II. Other problems may require direct treatment of sums of type I_r with $r \geqslant 3$ (see for instance the beautiful recent work of N. Pitt [36]). One might expect that this involves the theory of automorphic forms on GL_r.

Remark 1.18. In [14], the first and third authors obtained bounds which could be stronger than (1.6) and (1.7), in particular in ranges of M, N which are shorter than the Polya-Vinogradov range $p^{1/2}$ (see [14, Prop. 1.2 and Thm 1.4]). These bounds were established only for very special weights associated to rank one sheaves (additive characters of specific rational functions). It is quite conceivable that these results remain valid for more general trace weights, and we hope to come back to this question in a later work. From our current level of understanding at least, it seems that, instead of the Fourier-Möbius group (or in addition to it), we would need to involve more precise information on the underlying sheaf, for example concerning fine details of its ramification behavior, and/or its geometric monodromy group.

1.6. **Acknowledgments.** Part of this work was done during the 60th birthday conference of Roger Heath-Brown at Oxford. We would also like to thank the organizers, Tim Browning, David Ellwood and Jonathan Pila for this rich and pleasant week. Roger Heath-Brown has obtained fundamental results in the theory of the primes; it is no surprize that, once more, the celebrated "Heath-Brown identity" makes a crucial appearance in the present work.

This paper has benefited from discussions and comments from Jean Bourgain, Satadal Ganguly, Paul Nelson, Richard Pink, Ngô Bao Châu, Peter Sarnak, Akshay Venkatesh, and Zhiwei Yun and it is a pleasure to thank them for their input.

2. Algebraic twists of Eisenstein series and sums of type I_2

In this section, we will prove Theorem 1.14 and Theorem 1.15 simultaneously. Indeed, the two results are very closely related, as we will first clarify.

Let $M, N, X \geqslant 1$ with $X/4 \leqslant MN \leqslant 4X$. Let $t \in \mathbf{R}$ be given, as well as three smooth functions U, V, W satisfying (1.1) with respective parameters Q_U, Q_V, Q_W. We package these parameters by denoting

$$\boldsymbol{P} = (U, V, W, M, N, X),$$

and we denote

$$\mathcal{S}_{\boldsymbol{P}}(it, K) = \sum_{m,n} K(mn)\Bigl(\frac{m}{n}\Bigr)^{it} U\Bigl(\frac{m}{M}\Bigr) V\Bigl(\frac{n}{N}\Bigr) W\Bigl(\frac{mn}{X}\Bigr),$$

which is the sum involved in Theorem 1.15.

We start with a lemma relating the sums of type $\mathcal{S}_{V,X}(\cdot, K)$ and $\mathcal{S}_{\boldsymbol{P}}(\cdot, K)$.

Lemma 2.1. *We adopt the above notations and for $s \in \mathbb{C}$ and $x > 0$, let*

$$W_s(x) = W(x)x^{-s}.$$

(1) *For every $\varepsilon > 0$, there exists $C = C(\varepsilon)$, such that we have*

$$\mathcal{S}_{\boldsymbol{P}}(it, K) \ll_{\varepsilon} (Q_U + Q_V)^C + \iint_{|t_1|, |t_2| \leqslant (MN)^{\varepsilon}} |\mathcal{S}_{W_{t_1}, X}(it_2 + it, K)| dt_1 dt_2.$$

(2) *For every $\varepsilon > 0$, one has*

$$\mathcal{S}_{V,X}(it, K) \ll_{\varepsilon} X^{\varepsilon} \max_{\boldsymbol{P} = (U_1, V_1, V, M, N, X)} |\mathcal{S}_{\boldsymbol{P}}(it, K)|,$$

where $\boldsymbol{P}$ runs over parameters (U_1, V_1, V, M, N, X) as above with $Q_{U_1} = Q_{V_1} = 1$.

Proof. (1) Denote by $\hat{U}$ and $\hat{V}$ the Mellin transforms of the smooth functions U and V. These are entire functions, which satisfy

$$\hat{U}(s), \hat{V}(s) \ll \Bigl(\frac{Q_U + Q_V}{1 + |s|}\Bigr)^k, \tag{2.1}$$

for any $k \geqslant 0$, where the implicit constants depend on k, $\mathfrak{Re}\, s$ and the implicit constants in (1.1).

We then have

$$\mathcal{S}_{\boldsymbol{P}}(it, K) = \frac{1}{(2i\pi)^2} \int_{(0)} \int_{(0)} \hat{U}(u)\hat{V}(v)\mathcal{T}_W(u, v) N^u M^v du dv$$

by Mellin inversion, where

$$\mathcal{T}_W(u, v) = \sum_{m,n \geqslant 1} K(mn) m^{it-u} n^{-it-v} W\Bigl(\frac{mn}{X}\Bigr).$$

This sum can be expressed as a twist of Eisenstein series (1.5), namely

$$\mathcal{T}_W(u,v) = X^{-\theta_1}\mathcal{S}_{W_{\theta_1},X}(\theta_2+it,K),$$

where

$$\theta_1 = \frac{u+v}{2}, \qquad \theta_2 = \frac{-u+v}{2}.$$

Thus, by a change of variable, we get

$$\mathcal{S}_{\boldsymbol{P}}(it,K) = \frac{2}{(2i\pi)^2}\int_{(0)}\int_{(0)} \hat{U}(\theta_1-\theta_2)\hat{V}(\theta_1+\theta_2)\Bigl(\frac{M}{N}\Bigr)^{\theta_2}\Bigl(\frac{MN}{X}\Bigr)^{\theta_1}\mathcal{S}_{W_{\theta_1},X}(\theta_2+it,K)d\theta_2 d\theta_1.$$

The function W_{θ_1} is smooth and compactly supported on $[1/2,2]$. For $\mathfrak{Re}\,\theta_1=0$, it satisfies (1.1) with parameter

$$Q(\theta_1) \ll Q_W + |\theta_1|,$$

where the implicit constant is absolute.

Using (2.1) for k large enough, and the trivial bound

$$\mathcal{S}_{V,X}(it,K) \ll X(\log X),$$

the contribution to this double integral of the region where $|\theta_1| \geqslant (MN)^{\varepsilon}$ or $|\theta_2| \geqslant (MN)^{\varepsilon}$ is

$$\ll (Q_U+Q_V)^C$$

for some $C \geqslant 0$ depending only on ε, which concludes the proof.

(2) By a dyadic partition of unity (using Lemma 4.3 below), and taking into account the support condition, we can decompose $\mathcal{S}_{V,X}(it,K)$ into $O(\log X)$ sums of the shape

$$\mathcal{S}_{\boldsymbol{P}}(it,K)$$

where

$$\boldsymbol{P} = (U_1,V_1,V,M,N,X)$$

with $X/4 \leqslant MN \leqslant 4X$, and furthermore the functions U_1, V_1 satisfy condition (1.1) with parameters $Q_{U_1}=Q_{V_1}=1$. The result is then immediate. □

2.1. **A simple bound.** We start with the following simple "convexity" bound for the Eisenstein twists $\mathcal{S}_{V,X}(it,K)$, which is useful for $X \geqslant p$, and will indeed imply both Theorem 1.15 and Theorem 1.14 for $X \geqslant p^{5/4+\varepsilon}$.

Lemma 2.2. *With the notation and assumptions of Theorem 1.14, we have for any $\varepsilon>0$,*

$$\mathcal{S}_{V,X}(it,K) \ll \bigl(pQX(1+|t|)\bigr)^{\varepsilon}(1+|t|)^{1/2}QX\Bigl(\frac{1}{p}+\frac{p}{X}\Bigr)^{1/2}, \tag{2.2}$$

where the implicit constant depends on ε and polynomially on $\mathrm{cond}(\mathcal{F})$.

Proof. This is relatively standard, so we will be fairly brief: the idea is to use periodicity of K and to represent it in terms of Dirichlet characters, reducing then to easy estimates for moments of Dirichlet L-functions.

First of all, the contribution to $\mathcal{S}_{V,X}(it,K)$ of the integers n divisible by p is

$$\sum_{n\equiv 0\,(\mathrm{mod}\,p)} K(0)d_{it}(n)V\Bigl(\frac{n}{X}\Bigr) \ll_{\mathrm{cond}(\mathcal{F})} p^{-1}X\log X.$$

Next, for $(n,p)=1$, we can write

$$K(n) = \frac{1}{(p-1)^{1/2}}\sum_{\chi}\tilde{K}(\chi)\chi(n)$$

where χ runs over the Dirichlet characters modulo p and

$$\tilde{K}(\chi) = \frac{1}{(p-1)^{1/2}} \sum_{m \in \mathbf{F}_p^\times} K(m)\overline{\chi}(m)$$

is the finite-field Mellin transform of K. Thus we get

$$\sum_{(n,p)=1} K(n) d_{it}(n) V\Big(\frac{n}{X}\Big) = \frac{1}{(p-1)^{1/2}} \sum_{\chi} \tilde{K}(\chi) \sum_{n} \chi(n) d_{it}(n) V\Big(\frac{n}{X}\Big).$$

The contribution of the trivial character χ_0 to this sum is estimated by

$$\frac{1}{(p-1)^{1/2}} \tilde{K}(\chi_0) \sum_{n} \chi_0(n) d_{it}(n) V\Big(\frac{n}{X}\Big) \ll_{\mathrm{cond}(\mathcal{F})} p^{-1/2} X \log X$$

(indeed, since K is a trace weight, we have

$$(p-1)^{1/2}\tilde{K}(\chi_0) = \sum_{m \in \mathbf{F}_p} K(m) - K(0) = p^{1/2}\hat{K}(0) - K(0) \ll_{\mathrm{cond}(\mathcal{F})} p^{1/2},$$

where $\hat{K}$ denotes the unitarily normalized Fourier transform of K, so that $-\hat{K}$ is also an isotypic trace weight, associated to a sheaf with conductor bounded in terms of that of $\mathcal{F}$ only, by the properties of the Fourier transform of ℓ-adic sheaves, as explained in [12, §1.4, Prop. 8.2].)

For χ non-trivial, denoting by $\hat{V}(s)$ the Mellin transform of V, we have

$$\sum_{\chi \neq \chi_0} \tilde{K}(\chi) \sum_{n} \chi(n) d_{it}(n) V\Big(\frac{n}{X}\Big) = \frac{1}{2i\pi} \int_{\mathfrak{Re}\, s = 1/2} \sum_{\chi \neq \chi_0} \tilde{K}(\chi) L(\chi, s+it) L(\chi, s-it) \tilde{V}(s) X^s ds,$$

by a standard application of Mellin inversion and a contour shift.

From (1.1), we get

$$\tilde{V}(s) \ll_j \Big(\frac{Q}{1+|s|}\Big)^j$$

for all $j \geqslant 0$, where the implicit constant depends on j. Now, for any fixed $\varepsilon > 0$, let $S = Q^{1+\varepsilon}$, and split the s-integral into

$$\frac{1}{2i\pi} \int_{\mathfrak{Re}\, s = 1/2} \cdots = \frac{1}{2i\pi} \int_{\substack{\mathfrak{Re}\, s = 1/2 \\ |\mathfrak{Im}\, s| \leqslant S}} \cdots + \frac{1}{2i\pi} \int_{\substack{\mathfrak{Re}\, s = 1/2 \\ |\mathfrak{Im}\, s| > S}} \cdots = I_1 + I_2,$$

say. To handle I_1, we apply Cauchy's inequality to obtain

$$I_1^2 \ll \Big\{ XS \sum_{\chi} |\tilde{K}(\chi)|^2 \Big\} \times \int_{\substack{\mathfrak{Re}\, s = 1/2 \\ |\mathfrak{Im}\,(s)| \leqslant S}} \sum_{\chi \neq \chi_0} |L(\chi, s-it) L(\chi, s+it)|^2 |ds|.$$

By the Parseval identity, the first factor on the right-hand side is $\ll pXS$. For the second factor, we apply the approximate functional equation to bound the product $L(\chi, s-it)L(\chi, s+it)$ by sums of the shape

$$\sum_{n} \frac{\chi(n) d_{it}(n)}{n^s} W\Big(\frac{n}{N}\Big)$$

for W rapidly decreasing and $N \ll p(1 + S + |t|)$ (see, e.g., [26, Th. 5.3]). By a hybrid-large sieve estimate (see [35, Th. 6.4], compare with [26, Th. 7.34]), we can get the Lindelöf conjecture on average for the integral and sum, and therefore derive

$$\int_{\substack{\mathfrak{Re}\, s = 1/2 \\ |\mathfrak{Im}\,(s)| \leqslant S}} \sum_{\chi \neq \chi_0} |L(\chi, s-it) L(\chi, s+it)|^2 |ds| \ll_\varepsilon (pS(1+|t|))^{1+\varepsilon}.$$

Using the rapid decay of $\tilde{V}(s)$, a similar bound holds for I_2, and this gives the desired result. □

For $X \geqslant p^{3/2}$, for instance, the bound in this lemma is stronger than the one claimed in Theorem 1.14. Using Lemma 2.1 (1), we then also deduce Theorem 1.15 in this range. We will therefore assume for the remainder of this section that $X \leqslant p^{3/2}$. Similarly, comparing the bounds of Theorems 1.14 and 1.15 with the trivial bounds

$$\mathcal{S}_{V,X}(it,K) \ll X\log X, \qquad \mathcal{S}_{\mathbf{P}}(it,K) \ll X\log X,$$

we may assume that $X \geqslant p^{3/4}$.

2.2. **Spectral theory and amplification.** The most important ingredient in the proof of Theorems 1.14 and 1.15, is the following proposition, which is proved with the methods of [12], based on the amplification method and Kuznetsov's formula. It is an averaged version of a bound for the amplified second moment of the sums $\mathcal{S}_{V,X}(it,K)$.

For $\tau \in \mathbf{R}$, $L \geqslant 1$ and $u \in \mathbf{C}$, let

$$B_{i\tau}(u) = \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} \operatorname{sign}(d_{i\tau}(\ell))d_u(\ell),$$

which is the amplifier (of length $2L$) adapted to the Eisenstein series $E(z,1/2+i\tau)$.

Lemma 2.3. *For any $\varepsilon > 0$ there exists $b = b(\varepsilon) \geqslant 0$ such that*

$$\int_{\mathbf{R}} \min(|t|^2, |t|^{-2-2b})|B_{i\tau}(it)\mathcal{S}_{V,X}(it,K)|^2 dt \ll_{\varepsilon} p^{\varepsilon}\big(pLXQ + p^{1/2}L^3XQ(X/p+Q)^2\big), \tag{2.3}$$

provided

$$LQ < p^{1/4}, \qquad 1 \leqslant L \leqslant X.$$

Proof. As in the cuspidal case in [12], we use the amplification method and the Kuznetsov formula, exploiting the fact that, for any given $\tau \in \mathbf{R}$, the Eisenstein series

$$\frac{1}{(p+1)^{1/2}}E(z,1/2+i\tau)$$

occurs in the continuous spectrum of Hecke eigenforms of level p. More precisely, we have the Fourier expansion

$$E(z,1/2+it) = y^{1/2+it} + \frac{\theta(1/2-it)}{\theta(1/2+it)}y^{1/2-it} + \frac{1}{\theta(1/2+it)}\sum_{n\neq 0} d_{it}(|n|)|n|^{-1/2}W_{it}(4\pi|n|y)e(nx),$$

where

$$\theta(s) = \pi^{-s}\Gamma(s)\zeta(2s),$$

and

$$W_{it}(y) = \frac{e^{-y/2}}{\Gamma(it+\frac{1}{2})}\int_0^{\infty} e^{-x}x^{it-1/2}\Big(1+\frac{x}{y}\Big)^{it-1/2}dx \tag{2.4}$$

denotes the Whittaker function (see for instance [25, (3.29)]).

We assume that the condition of the lemma are met. Using the notation of [12, Section 4] and taking there $P = Xp^{-1}$, we obtain as in [12, Prop. 4.1, (4.10)], the bound

$$\begin{aligned}\frac{1}{p+1}\int_{\mathbf{R}} \tilde{\phi}_{a,b}(t)\frac{1}{\cosh(\pi t)|\theta(1/2+it)|^2}|B_{i\tau}(it)|^2|\mathcal{S}_{V,X}(it,K)|^2 dt \\ \leqslant M(L) - 2\sum_{k>a-b}\dot{\phi}(k)(k-1)M(L;k) \ll p^{\varepsilon}\Big(LXQ + \frac{L^3X}{p^{1/2}}Q\Big(\frac{X}{p}+Q\Big)^2\Big)\end{aligned} \tag{2.5}$$

for any $\varepsilon > 0$, where $2 \leqslant b < a$ are odd integers depending on ε and $\tilde{\phi}(t) = \tilde{\phi}_{a,b}(t)$ denotes a positive function such that

$$\tilde{\phi}(t) \asymp_{a,b} (1+|t|)^{-2b-2}.$$

We then obtain the desired average estimate from this, using Stirling's formula and the bound

$$\zeta(1+2it) \ll \frac{1}{|t|} + \log(2+|t|).$$

□

In order to apply this, we need some lower bound for the amplifier $B_{i\tau}(it)$. The following lemma gets a suitable bound for t close enough to τ.

Lemma 2.4. *For L large enough, we have*

$$B_{i\tau}(it) \gg \frac{L}{\log^3 L},$$

uniformly for t and $\tau \in \mathbf{R}$ satisfying

$$|t-\tau| \ll \frac{1}{\log^4 L}, \qquad \textit{and} \qquad |\tau| \leqslant L^{\frac{1}{3}}.$$

Proof. We observe first that for any prime $\ell \leqslant 2L$ and $|t-\tau| \ll \log^{-4} L$, we have

$$\begin{aligned}|B_{i\tau}(it) - B_{i\tau}(i\tau)| &\leqslant \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} |d_{i\tau}(\ell) - d_{it}(\ell)| \\ &= 2 \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} |\cos(\tau \log \ell) - \cos(t \log \ell)| \\ &\leqslant 2|t-\tau| \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} \log \ell \ll \frac{L}{\log^4 L},\end{aligned}$$

and hence it suffices to prove the lower bound for $t = \tau$.

Furthermore, we may clearly assume that $\tau > 0$ (by parity) and that $L \geqslant 3$. We then have

$$B_{i\tau}(i\tau) = \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} \operatorname{sign}(d_{i\tau}(\ell)) d_{i\tau}(\ell) = \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} |d_{i\tau}(\ell)| = 2 \sum_{\substack{\ell \leqslant 2L \\ \ell \text{ prime}}} |\cos(\tau \log \ell)|,$$

and since $|\cos(\tau \log \ell)| \leqslant 1$ it is enough to prove that

$$\sum_{\ell \sim L} \cos(\tau \log \ell)^2 \gg \frac{L}{\log^3 L}$$

(where ℓ ranges over primes $L < \ell \leqslant 2L$) under the assumption of the lemma. We do this by finding suitable sub-intervals where $\tau \log \ell$ is sufficiently far away from $\pi/2$ modulo $\pi\mathbf{Z}$.

Consider the function

$$g(x) = \tau \log x$$

for $x \in [L, 2L]$. It is non-decreasing and satisfies

$$g(2L) - g(L) = \tau \log 2, \; g'(x) \in \Big[\frac{\tau}{2L}, \frac{\tau}{L}\Big] \text{ for } x \in [L, 2L].$$

In particular, if $\tau \log 2 \geqslant 2\pi$, the preimage $g^{-1}([-\pi/4, \pi/4] + \pi\mathbf{Z})$ intersects $[L, 2L]$ in $\gg 1/\tau$ intervals of length $\geqslant \frac{\pi L}{2\tau}$. From Huxley's Theorem on primes in short intervals (see, e.g., [26, Th. 10.4, Th. 5] and note that any of the variants of [26, Th. 10.5], going back to Hoheisel, would be enough) the number of primes in any such interval is $\gg \frac{L}{\tau \log L}$, provided $\tau \leqslant L^{5/12-\varepsilon}$ for some fixed $\varepsilon > 0$ and L is large enough. Therefore (taking $\varepsilon = 1/12$) we obtain

$$|B_{i\tau}(i\tau)| \gg \frac{L}{\log L}$$

provided $2\pi/\log 2 \leqslant \tau \leqslant L^{1/3}$.

At the other extreme, if $0 \leqslant \tau \leqslant \frac{1}{100 \log L}$, we have

$$\cos^2(\tau \log \ell) \geqslant \cos^2\Bigl(\frac{1}{50}\Bigr)$$

for every $L \leqslant \ell \leqslant 2L$, and hence $B_{i\tau}(i\tau) \gg L/\log L$ also in that case.

Suppose now that

$$\frac{1}{100 \log L} \leqslant \tau \leqslant \frac{2\pi}{\log 2}.$$

In that case, $g([L, 2L])$ is an interval of length at least $1/(200 \log L)$. It is then easy to see (the worse case is when the interval is symmetric around $\pi/2 + k\pi$ for some integer k) that there exists $x_0 \in [L, 2L]$ such that

$$\cos^2(\tau \log(x_0)) \geqslant \frac{1}{2 \cdot 400^2 \log^2 L}.$$

Using again Huxley's Theorem, we know that the interval $[L, 2L] \cap [x_0 - L^{2/3}, x_0 + L^{2/3}]$ contains at least $\gg L/\log L$ primes, and since

$$|\cos^2(\tau \log(\ell)) - \cos^2(\tau \log(x_0))| \ll |\log(\ell/x_0)| \ll L^{-1/3}$$

for these primes, we have

$$\cos^2(\tau \log(\ell)) \gg \log^{-2} L,$$

and therefore

$$B_{i\tau}(i\tau) \gg \frac{L}{\log^3 L}$$

in that last case, which concludes the proof. □

This lemma and the average bound (2.3) allow us to deduce a first good upper-bound for the twists of Eisenstein series, averaged in rather short intervals. It will be convenient for later purposes to introduce the notation

$$I(\tau, p) = \{t \in \mathbf{R} \mid |t - \tau| \leqslant \log^{-4} p\}, \quad \mathcal{M}(\tau, p) = \max_{t \in I(\tau,p)} |\mathcal{S}_{V,X}(it, K)|, \tag{2.6}$$

$$M(Q, X) = QX\Bigl(1 + \frac{p}{X}\Bigr)^{1/2} p^{-1/8}, \tag{2.7}$$

so that, for instance, Theorem 1.14 claims that

$$\mathcal{S}_{V,X}(it, K) \ll p^{\varepsilon}(1 + |t|)^A M(Q, X)$$

for any $\varepsilon > 0$ and $A \geqslant 1$ depending on ε.

Our next result is:

Proposition 2.5. *For any $\tau \in \mathbf{R}$, we have*

$$\int_{I(\tau,p)} \min(|t|^2, 1)|\mathcal{S}_{V,X}(it, K)|^2 dt \ll_{\varepsilon} p^{\varepsilon}(1 + |\tau|)^B M(Q, X)^2, \tag{2.8}$$

for any $\varepsilon > 0$ and some $B \geqslant 1$ which depends only on ε.

Proof. Let

$$L = \frac{p^{1/4}}{X/p + Q}$$

as in [12]. If either $L \ll 1$, or $|\tau| > L^{1/3}$, the trivial bound

$$\mathcal{S}_{V,X}(i\tau, K) \ll X \log X$$

or the convexity bound (2.2) yield stronger results than (2.8) (since $B \geqslant 1$). Otherwise, combining Lemma 2.4 with (2.3), we obtain

$$\begin{aligned}\int_{|t-\tau|\leqslant \log^{-4} p} \min(|t|^2, |t|^{-2-2b})|\mathcal{S}_{V,X}(it,K)|^2 dt &\ll_\varepsilon p^\varepsilon\Big(\frac{pXQ}{L} + p^{1/2}XQL\Big(\frac{X}{p}+Q\Big)^2\Big)\\ &\ll p^\varepsilon Q^2X^2\Big(1+\frac{p}{X}\Big)p^{-1/4},\end{aligned}$$

for some B depending on ε, as desired. □

The remaining objective is to derive a pointwise bound for $\mathcal{S}_{V,X}(it, K)$, and to do so we must relax the zero of order 2 of the weight $\min(|t|^2, 1)$ at the origin (for similar issues with estimates of L-functions, see e.g. [1]; we could use similar methods, but at the expense of expressing our sums in terms of L-functions, and instead we use them directly, and resort to an iterative argument.)

The basic mechanism is the following consequence of the proposition:

Corollary 2.6. *For any $\tau \in \mathbf{R}$ and $\varepsilon > 0$, with notation as above, we have*

$$\int_{I(\tau,p)} |\mathcal{S}_{V,X}(it,K)|dt \ll_\varepsilon \begin{cases} p^\varepsilon(1+|\tau|)^B M(Q,X) & \text{if } |\tau| \geqslant 1,\\ p^\varepsilon \mathcal{M}(\tau,p)^{1/3}M(Q,X)^{2/3} & \text{if } |\tau| \leqslant 1.\end{cases}$$

Proof. If $|\tau| \geqslant 1$, we just apply the Cauchy-Schwarz inequality to get

$$\begin{aligned}\int_{I(\tau,p)} |\mathcal{S}_{V,X}(it,K)|dt &\leqslant \Big(\int_{I(\tau,p)} |\mathcal{S}_{V,X}(it,K)|^2dt\Big)^{1/2}\Big(\int_{I(\tau,p)} dt\Big)^{1/2}\\ &\ll \Big(\int_{I(\tau,p)} \min(|t|^2,1)|\mathcal{S}_{V,X}(it,K)|^2 dt\Big)^{1/2} \ll p^\varepsilon(1+|\tau|)^B M(Q,X)\end{aligned}$$

by Proposition 2.5.

Now assume $|\tau| < 1$. Let $0 < \alpha < 1/3$ be some parameter. By Hölder's inequality, we get

$$\begin{aligned}\int_{I(\tau,p)} |\mathcal{S}_{V,X}(it,K)|dt &\leqslant \mathcal{M}(\tau,p)^{1-2\alpha}\int_{I(\tau,p)} |\mathcal{S}_{V,X}(it,K)|^{2\alpha}dt\\ &\leqslant \mathcal{M}(\tau,p)^{1-2\alpha}\Big(\int_{I(\tau,p)} |t|^2|\mathcal{S}_{V,X}(it,K)|^2dt\Big)^\alpha\Big(\int_0^2 |t|^{-2\alpha/(1-\alpha)}dt\Big)^{1-\alpha}\\ &\ll_{\varepsilon,\alpha} p^\varepsilon \mathcal{M}(\tau,p)^{1-2\alpha}M(Q,X)^{2\alpha}\\ &\ll_{\varepsilon,\alpha} p^\varepsilon \mathcal{M}(\tau,p)^{1-2\alpha}M(Q,X)^{2/3},\end{aligned}$$

(since $2\alpha/(1-\alpha) < 1$ and $M(Q,X) \geqslant 1$). By the trivial bound, we have

$$\mathcal{M}(\tau,p)^{1-2\alpha} \ll \mathcal{M}(\tau,p)^{1/3}(X\log X)^{1-2\alpha-1/3},$$

and we conclude by taking $\alpha = 1/3 - \varepsilon$. □

2.3. **An iterative bound.** The following lemma establishes an iterative bound for Eisenstein twists.

Lemma 2.7. *Assume that $\beta > 0$ is such that*

$$\mathcal{S}_{V,X}(it, K) \ll p^{\varepsilon}(1+|t|)^A X^{\beta} M(Q,X)^{1-\beta} \tag{2.9}$$

for $X \leqslant p^{3/2}$, any $\varepsilon > 0$, and some $A \geqslant 1$ depending on ε. Then we have

$$\mathcal{S}_{V,X}(it, K) \ll p^{\varepsilon}(1+|t|)^{A_1} X^{\beta/3} M(Q,X)^{1-\beta/3} \tag{2.10}$$

$$\mathcal{S}_{\boldsymbol{P}}(it, K) \ll p^{\varepsilon}(Q_U+Q_V)^B(1+|t|)^{A_1} X^{\beta/3} M(Q_W,X)^{1-\beta/3} \tag{2.11}$$

for A_1, $B \geqslant 1$ depending on ε.

Proof. Using Lemma 2.1 (1), we first use the assumption to estimate $\mathcal{S}_{\boldsymbol{P}}(it, K)$. For each t_1, we split the integral over $|t_2| \leqslant p^{\varepsilon}$ into $\ll p^{\varepsilon}$ integrals over intervals of length $\log^{-4} p$. For an interval I with center at τ with $|\tau| \leqslant 1$, the integral is bounded by

$$\ll p^{\varepsilon}\mathcal{M}^{1/3} M(Q_W + |t_1|, X)^{2/3}$$

by Corollary 2.6, where

$$\mathcal{M} = \max_{t\in I} |\mathcal{S}_{W_{t_1},X}(it, K)| \ll p^{\varepsilon} X^{\beta} M(Q_W + |t_1|, X)$$

by (2.9) applied to W_{t_1}. Thus each such integral is

$$\ll p^{\varepsilon} X^{\beta/3} M(Q_W + |t_1|, X)^{1-\beta/3}.$$

For intervals centered at τ with $|\tau| \geqslant 1$, we obtain the bound $\ll p^{\varepsilon}(1+|\tau|)^A M(Q_W+|t_1|)$, which is better, and integrating over $|t_1| \leqslant p^{\varepsilon}$, we get (2.11) (note that $Q \mapsto M(Q,X)$ is linear).

Now, applying Lemma 2.1 (2), we immediately deduce (2.10). □

We are now done: for $X \leqslant p^{3/2}$, we can start applying this lemma with $\beta = 1$ by the trivial bound $\mathcal{S}_{V,X}(it, K) \ll X \log X$. We deduce that, for any integer $k \geqslant 1$, we have

$$\mathcal{S}_{V,X}(it, K) \ll p^{\varepsilon}(1+|t|)^A X^{3^{-k}} M(Q,X)^{1-3^{-k}}$$
$$\mathcal{S}_{\boldsymbol{P}}(it, K) \ll p^{\varepsilon}(Q_U+Q_V)^B(1+|t|)^A X^{3^{-k}} M(Q_W,X)^{1-3^{-k}}.$$

Since

$$\begin{aligned} X^{3^{-k}} M(Q,X)^{1-3^{-k}} &= XQ^{1-3^{-k}}\Big(1+\frac{p}{X}\Big)^{(1-3^{-k})/2} p^{-(1-3^{-k})/8} \\ &\leqslant XQ(1+p/X)^{1/2} p^{-1/8} p^{3^{-k}/8}, \end{aligned}$$

we therefore obtain Theorems 1.14 and 1.15 by taking k large enough.

3. Estimating sums of type *II*

In this section we prove Theorem 1.16. We will leave the proof of the simpler bound (1.7) to the reader, and consider (1.6), proceeding along classical lines. Denoting

$$T = \sum_{\substack{m,n \\ (m,p)=1}}\sum \alpha_m \beta_n K(mn)$$

the bilinear form to estimate, we apply Cauchy's inequality and deduce that

$$|T|^2 \leqslant \|\beta\|^2 \sum_{\substack{M/2 \leqslant m_1,m_2 \leqslant 2M \\ p\nmid m_1m_2}} \overline{\alpha_{m_1}}\alpha_{m_2} \sum_{N/2\leqslant n \leqslant 2N} \overline{K(m_1n)}K(m_2n). \tag{3.1}$$

The inner correlation coefficients are then treated by completion (i.e., by the Polya-Vinogradov method), which gives

$$(3.2)\qquad \sum_{N/2\leqslant n\leqslant 2N}\overline{K(m_1n)}K(m_2n)\ll \frac{N}{p}|\mathcal{C}(m_1,m_2,0,K)|+\sum_{0<|h|\leqslant p/2}\min\Bigl(\frac{1}{|h|},\frac{N}{p}\Bigr)|\mathcal{C}(m_1,m_2,h,K)|$$

where

$$\mathcal{C}(m_1,m_2,h,K)=\sum_{z\in\mathbf{F}_p}\overline{K(m_1z)}K(m_2z)e\Bigl(\frac{hz}{p}\Bigr),$$

a sum which satisfies the relation

$$\mathcal{C}(m_1,m_2,h,K)=\mathcal{C}(m_1/m_2,1,h/m_2,K).$$

For a trace weight, we have the trivial bound

$$|\mathcal{C}(m_1,m_2,h,K)|\leqslant \mathrm{cond}(\mathcal{F})^2p,$$

but this is not sharp in most cases. In fact, the crucial point is to show that for most parameters (m_1,m_2,h), we have a better estimate with square-root cancellation. We provide such a result in Theorem 6.3 in Section 6, building on our earlier work in [12].

Proposition 3.1 (Paucity of large correlations)**.** *Let K be an irreducible trace weight modulo p which is not p-exceptional, associated to the sheaf $\mathcal{F}$. Then there exists $C\geqslant 1$, $D\geqslant 0$, depending only polynomially on* $\mathrm{cond}(\mathcal{F})$*, such that*

$$|\mathcal{C}(m,1,h,K)|\leqslant Cp^{1/2}$$

for every pair $(m,h)\in\mathbf{F}_p^\times\times\mathbf{F}_p$ except for those in a set of pairs of cardinal at most D.

After inserting (3.2) in (3.1), the contribution of all triples (m_1,m_2,h) for which

$$|\mathcal{C}(m_1,m_2,h,K)|\leqslant Cp^{1/2}$$

is at most

$$\ll\|\alpha\|^2\|\beta\|^2\Bigl(\frac{MN}{p^{1/2}}+Mp^{1/2}\log p\Bigr).$$

For the remaining triples, we sum over m_1 first. For each m_1, the proposition shows that the possible $(m_1/m_2,h/m_1)$ that can occur lie, modulo p, in a finite set $\mathcal{E}$ of size bounded in terms of the conductor of $\mathcal{F}$ only, i.e., m_2 modulo p and h are determined by m_1 up to a finite number of possibilities. We use the trivial bounds

$$|\mathcal{C}(m_1,m_2,h,K)|\leqslant \mathrm{cond}(\mathcal{F})^2p,\qquad \min\Bigl(\frac{1}{|h|},\frac{N}{p}\Bigr)\leqslant\frac{N}{p},$$

and obtain that the contribution of these terms to the right-hand side of (3.1) is

$$\begin{aligned}&\ll\|\beta\|^2N\sum_{(t,h)\in\mathcal{E}}\sum_{m_1}\sum_{\substack{m_1,m_2\\ m_2\equiv tm_1\,(\mathrm{mod}\,p)}}|\alpha_{m_1}||\alpha_{m_2}|\\ &\ll\|\beta\|^2N\sum_{(t,h)\in\mathcal{E}}\sum_{m_1}\sum_{\substack{m_1,m_2\\ m_2\equiv tm_1\,(\mathrm{mod}\,p)}}(|\alpha_{m_1}|^2+|\alpha_{m_2}|^2)\\ &\ll N\Bigl(1+\frac{M}{p}\Bigr)\|\beta\|^2\|\alpha\|^2,\end{aligned}$$

where the implicit constant depends only (polynomially) on $\mathrm{cond}(\mathcal{F})$.

Combining the two, we get

$$T \ll \|\alpha\|\|\beta\|(MN)^{1/2}\Bigl(\frac{1}{p^{1/4}}+\frac{1}{M^{1/2}}+\frac{p^{1/4}\log^{1/2}p}{N^{1/2}}\Bigr),$$

where the implicit constant depends only on the conductor of $\mathcal{F}$. This completes the proof of Theorem 1.16.

4. Sums over primes

We now finally prove Theorem 1.5, our main result on sums over primes.

4.1. **Smooth sums.** We start with the smooth version (1.2). Clearly, it is enough to estimate the sum

$$\mathcal{S}_{V,X}(\Lambda,K)=\sum_n \Lambda(n)K(n)V\Bigl(\frac{n}{X}\Bigr),$$

and we begin by recalling two lemmas. The first one is Heath-Brown's identity for the von Mangoldt function [22]:

Lemma 4.1 (Heath-Brown). *For any integer $J\geqslant 1$ and $n<2X$, we have*

$$\Lambda(n)=\sum_{j=1}^{J}(-1)^j\binom{J}{j}\sum_{m_1,\cdots,m_j\leqslant Z}\mu(m_1)\cdots\mu(m_j)\sum_{m_1\cdots m_jn_1\cdots n_j=n}\log n_1,$$

where $Z=X^{1/J}$.

Remark 4.2. Using instead the analogous formula

$$\mu(n)=\sum_{j=1}^{J}(-1)^j\binom{J}{j}\sum_{m_1,\cdots,m_j\leqslant Z}\mu(m_1)\cdots\mu(m_j)\sum_{m_1\cdots m_jn_1\cdots n_j=n}1,$$

for the Möbius function (valid under the same conditions), one proves Theorem 1.7 using exactly the same arguments, so we will not say more about the proof of that result.

The second lemma provides a smooth partition of unity (see, e.g., [11, Lemma 2]).

Lemma 4.3. *There exists a sequence $(V_l)_{l\geqslant 0}$ of smooth functions on $[0,+\infty[$ such that*

- *For any l, V_l is supported in $]2^{l-1},2^{l+1}[$;*
- *For any $k,l\geqslant 0$, we have*

$$x^kV_l^{(k)}(x)\ll_k 1,$$

where the implicit constant depends only on k;

- *For any $x\geqslant 1$,*

$$\sum_{l\geqslant 0}V_l(x)=1.$$

Fix some $J\geqslant 2$. Applying these two lemmas, we see that $\mathcal{S}_{V,X}(\Lambda,K)$ decomposes into a linear combination, with coefficients bounded by $O_J(\log X)$, of $O(\log^{2J}X)$ sums of the shape

$$\Sigma(\boldsymbol{M},\boldsymbol{N})=\sum\sum_{m_1,\cdots,m_J}\alpha_1(m_1)\alpha_2(m_2)\cdots\alpha_J(m_J)\times\sum\sum_{n_1,\cdots,n_J}V_1(n_1)\cdots V_J(n_J)V\Bigl(\frac{m_1\cdots m_Jn_1\cdots n_J}{X}\Bigr)K(m_1\cdots m_Jn_1\cdots n_J)\qquad(4.1)$$

where

- $\boldsymbol{M} = (M_1, \cdots, M_J)$, $\boldsymbol{N} = (N_1, \cdots, N_J)$ are J-uples of parameters in $[1/2, 2X]^{2J}$ which satisfy
$$N_1 \geqslant N_2 \geqslant \cdots \geqslant N_J, \qquad M_i \leqslant X^{1/J}, \qquad M_1 \cdots M_J N_1 \cdots N_J \asymp_J X;$$
- the arithmetic functions $m \mapsto \alpha_i(m)$ are bounded and supported in $[M_i/2, 2M_i]$;
- the smooth functions $V_i(x)$ are compactly supported in $[N_i/2, 2N_i]$, and their derivatives satisfy
$$y^k V_i^{(k)}(y) \ll 1,$$
where the implicit constants depend only on k.

We will state different bounds for $\Sigma(\boldsymbol{M}, \boldsymbol{N})$, depending on the relative sizes of the parameters, and then optimize the result.

First of all, from (1.7) and integration by parts, we get the bound

$$\Sigma(\boldsymbol{M}, \boldsymbol{N}) \ll (pQ)^{\varepsilon} QX \Big(\frac{1}{p^{1/2}} + \frac{p^{1/2}}{N_1} \Big) \tag{4.2}$$

for any $\varepsilon > 0$, the implicit constant depending on ε.

Next, for $J \geqslant 2$, we obtain, by Theorem 1.15 applied to n_1, n_2 and trivial summation over the remaining variables, the bound

$$\Sigma(\boldsymbol{M}, \boldsymbol{N}) \ll (pQ)^{\varepsilon} QX \Big(1 + \frac{p}{N_1 N_2} \Big)^{1/2} p^{-1/8}. \tag{4.3}$$

Finally from Theorem 1.16 with an integration by parts, we have the bound

$$\Sigma(\boldsymbol{M}, \boldsymbol{N}) \ll (pQ)^{\varepsilon} QX \Big(\frac{1}{p^{1/4}} + \frac{1}{M^{1/2}} + \frac{p^{1/4}}{(X/M)^{1/2}} \Big), \tag{4.4}$$

for any factorization

$$M_1 \cdots M_J N_1 \cdots N_J = M \times N$$

where M and N are products of some of the M_i and N_j.

Our goal is to choose the best of the three bounds above for each such configuration of the parameters $(\boldsymbol{M}, \boldsymbol{N})$. By taking logarithms (in base p), we readily see that the proof of (1.2) is reduced to the optimization problem of the next section.

4.2. An optimization problem.

We consider here the following optimization problem. We are given a real number $x > 0$ (we have in mind $x = \log X / \log p$), an integer $J \geqslant 3$, and parameters

$$(\boldsymbol{m}, \boldsymbol{n}) = (m_1, \cdots, m_J, n_1, \cdots, n_J) \in [0, x]^{2J}$$

such that

$$\sum_i m_i + \sum_j n_j = x, \qquad m_i \leqslant x/J, \qquad n_1 \geqslant n_2 \geqslant \cdots \geqslant n_J. \tag{4.5}$$

We want to estimate from below the quantity

$$\eta(\boldsymbol{m}, \boldsymbol{n}) = \max\Big\{ \min\Big(\frac{1}{2}, n_1 - \frac{1}{2} \Big), \max_{\sigma} \min\Big(\frac{1}{4}, \frac{\sigma}{2}, \frac{x - \sigma}{2} - \frac{1}{4} \Big), \ \frac{1}{8} - \max\Big(0, \frac{1}{2}(1 - (n_1 + n_2)) \Big) \Big\},$$

where σ ranges over all possible sub-sums of the m_i and n_j for $1 \leqslant i, j \leqslant J$, that is over the sums

$$\sigma = \sum_{i \in \mathcal{I}} m_i + \sum_{j \in \mathcal{J}} n_j$$

for $\mathcal{I}$, $\mathcal{J}$ ranging over all possible subsets of $\{1, \cdots, J\}$.

The number $\eta(\boldsymbol{m},\boldsymbol{n})$ represent the maximal power of p that we save over the trivial bound by using either (4.2) or (4.3) or (4.4) . The outcome of the discussion in the previous section is that, for $x=(\log X)/(\log p)$ and $J\geqslant 3$, we have

$$\Sigma_J(\boldsymbol{M},\boldsymbol{N})\ll (pQ)^{\varepsilon}QXp^{-\eta(\boldsymbol{m},\boldsymbol{n})}.$$

By Heath-Brown's identity, it follows that

$$\mathcal{S}_V(\Lambda,K)\ll (pQ)^{\varepsilon}QXp^{-\eta}$$

where

$$\eta=\min_{(\boldsymbol{m},\boldsymbol{n})}\eta(\boldsymbol{m},\boldsymbol{n}).$$

We will show:

Proposition 4.4. *Let $x>3/4$ be given. Provided J is large enough in terms of x, we have the inequality*

$$\eta(\boldsymbol{m},\boldsymbol{n})\geqslant\min\Bigl(\frac{1}{24},\frac{4x-3}{24}\Bigr).$$

Combining these lower-bounds with the above estimates, the proof of Theorem 1.5 is concluded, noting that $x\leqslant 1$ means that $X\leqslant p$, and that

$$Xp^{-(4x-3)/24}=X\Bigl(\frac{p}{X}\Bigr)^{1/6}p^{-1/24}.$$

Proof of Proposition 4.4. We have

$$\eta(\boldsymbol{m},\boldsymbol{n})\geqslant\max\Bigl\{\max_{\sigma}\min\Bigl(\frac{1}{4},\frac{\sigma}{2},\frac{x-\sigma}{2}-\frac{1}{4}\Bigr),\frac{1}{8}-\max(0,\frac{1}{2}(1-(n_1+n_2)))\Bigr\}. \tag{4.6}$$

Let δ be a parameter such that

$$0<\delta<\min\Bigl(\frac{4x-3}{12},\frac{x-1/2}{6},\frac{1}{4}\Bigr). \tag{4.7}$$

The interval

$$I_\delta=\Bigl[2\delta,x-\frac{1}{2}-2\delta\Bigr]$$

is then non-empty. If we can find a subsum σ such that $\sigma\in I_\delta$, we then deduce immediately that

$$\eta(\boldsymbol{m},\boldsymbol{n})\geqslant\max_{\sigma}\min\Bigl(\frac{1}{4},\frac{\sigma}{2},\frac{x}{2}-\frac{1}{4}-\frac{\sigma}{2}\Bigr)\geqslant\delta. \tag{4.8}$$

We now assume that such a subsum σ does *not* exist, and attempt to get a lower-bound on $\eta(\boldsymbol{m},\boldsymbol{n})$ using the second term in the maximum (4.6). First of all, we claim that, in that case, we have

$$\sum_{i\leqslant J}m_i<2\delta, \tag{4.9}$$

provided

$$\frac{1}{J}\leqslant x-\frac{1}{2}-4\delta=\mathrm{length}(I_\delta),$$

a condition which we assume from now on.

Indeed, if (4.9) were false, using the fact that $m_i\leqslant 1/J$ and that $1/J$ is then less than the length of the interval I_δ, we would be able to find some subsum σ (formed only with some m_i's) which is contained in I_δ, contradicting our current assumption.

From (4.5) and (4.9), we get in particular the inequality

$$\sum_j n_j\geqslant x-2\delta.$$

Since, under our assumption (4.7) on δ, we have

$$2\delta \leqslant x - \frac{1}{2} - 4\delta = \mathrm{length}(I_\delta),$$

this implies that

$$n_j \leqslant 2\delta$$

for any $j \geqslant 3$ (because otherwise, we would have

$$x - \frac{1}{2} - 2\delta \leqslant n_3 \leqslant n_2 \leqslant n_1$$

since $n_j \notin I_\delta$, and then, in view of (4.7), we would get

$$n_1 + n_2 + n_3 > 3x - \frac{3}{2} - 6\delta \geqslant x,$$

a contradiction). But now it follows that

$$\sum_{j\geqslant 3} n_j < 2\delta, \tag{4.10}$$

because otherwise, using $4\delta \leqslant x - 1/2 - 2\delta$, we could again obtain a subsum of the n_j's, $j \geqslant 3$, in I_δ.

Combining (4.9) and (4.10), we obtain

$$n_1 + n_2 \geqslant x - 4\delta$$

and hence

$$\frac{1}{8} - \max\Bigl(0, \frac{1}{2}(1 - (n_1 + n_2))\Bigr) \geqslant \min\Bigl(\frac{1}{8}, \frac{4x-3}{8} - 2\delta\Bigr).$$

Combining this with (4.8), it follows that for δ satisfying (4.7) and J large enough in terms of x and δ, we have

$$\eta(\boldsymbol{m}, \boldsymbol{n}) \geqslant \min\Bigl(\delta,\ \min\Bigl(\frac{1}{8}, \frac{4x-3}{8} - 2\delta\Bigr)\Bigr),$$

For $x > 3/4$, we take

$$\delta = \min\Bigl(\frac{4x-3}{24}, \frac{1}{24}\Bigr)$$

and Proposition 4.4 follows. □

4.3. **Sums over intervals.** We can now also easily deduce from (1.2) the estimate (1.3) for sums over primes in the interval $2 \leqslant q \leqslant X$ (below all sums over q are restricted to q prime). By a dyadic decomposition of the interval $[1, X]$, we are reduced to proving that

$$\sum_{X\leqslant q\leqslant 2X} K(q) \ll_{\eta,\mathrm{cond}(\mathcal{F})} X(1 + p/X)^{1/12} p^{-\eta/2} \tag{4.11}$$

for $X \geqslant 2$ and for any $\eta < 1/24$. Since the right-hand side of this bound increases with X, this is sufficient to conclude the proof of (1.3).

Let $\Delta < 1$ be some parameter and let V be a smooth function defined on $[0, +\infty[$ such that

$$\mathrm{supp}(V) \subset [1-\Delta, 2+\Delta], \qquad 0 \leqslant V \leqslant 1, \qquad V(x) = 1 \text{ for } 1 \leqslant x \leqslant 2,$$

and which satisfies

$$x^j V^{(j)}(x) \ll_j Q^j,$$

with $Q = \Delta^{-1}$.

By applying (1.2) to V, we get

$$\sum_{X\leqslant q\leqslant 2X} K(q) \ll X\Delta + \sum_{q} K(q)V\Big(\frac{q}{X}\Big)$$
$$\ll_{\eta,\mathrm{cond}(\mathcal{F})} X(\Delta + \Delta^{-1}(1+p/X)^{1/6}p^{-\eta})$$

for any $\eta < 1/24$.

If $X > p^{3/4}$, we can take

$$\Delta = (1+p/X)^{1/12}p^{-\eta/2} < 1$$

and we obtain (4.11). On the other hand, if $X \leqslant p^{3/4}$, the bound (4.11) is weaker than the trivial bound $2X$ for p large enough.

5. Applications

5.1. **Primes represented by a polynomial modulo p.** In this section we prove Corollaries 1.10 and 1.11.

For the former, we fix a non-constant polynomial $P \in \mathbf{Z}[X]$, and we consider a prime p such that P is non-constant modulo p.

For Corollary 1.10, (1), we are dealing with

$$\sum_{n\in\mathbf{F}_p} E(X;p,P(n)) = \sum_{n\in\mathbf{F}_p} \pi(X;p,P(n)) - \frac{1}{p-1} \sum_{\substack{n\in\mathbf{F}_p\\ P(n)\not\equiv 0 \,(\mathrm{mod}\, p)}} \pi(X).$$

We denote

$$N_P(x) = \sum_{\substack{n\in\mathbf{F}_p\\ P(n)=x}} 1 - 1$$

the "centered" number of representations of x as a value of P modulo p. The formula above allows us to write

$$\sum_{n\in\mathbf{F}_p} E(X;p,P(n)) = \sum_{q\leqslant X} N_P(q) + \sum_{q\leqslant X}\Big(1 - \frac{1}{p-1}|\{n\in\mathbf{F}_p \mid P(n)\neq 0\}|\Big)$$

(where q runs over primes, as before).

The second term of the previous expression is trivially bounded by $\ll p^{-1}X + 1$, since P has at most $\deg P$ zeros modulo p. Thus Corollary 1.10, (1) follows from Theorem 1.5 and from the fact – recalled in Section 6.2 below – that N_P is a trace function for an ℓ-adic sheaf with no exceptional Jordan-Hölder factor (i.e. no such factor is geometrically isomorphic to a tensor product of a Kummer sheaf and an Artin-Schreier sheaf).

For Corollary 1.10, (2), we write $\mathbf{1}_{P(\mathbf{F}_p)}$ for the characteristic function of the set $P(\mathbf{F}_p)$ of values of P modulo p, and we will denote $P^*(\mathbf{F}_p) = P(\mathbf{F}_p) - \{0\}$, the set of non-zero values of P modulo p. A reasoning similar to the previous one leads to

$$\sum_{a\in P(\mathbf{F}_p)} E(X;p,a) = \sum_{q\leqslant X} \mathbf{1}_{P(\mathbf{F}_p)}(q) - \frac{|P^*(\mathbf{F}_p)|}{p-1}\pi(X).$$

Applying Proposition 6.7 of Section 6.2, the first term on the right-hand side becomes

$$c_1\pi(X) + \sum_{2\leqslant i\leqslant k}\sum_{q\leqslant X} K_i(q) + O(p^{-1}X+1)$$

where the implicit constant depends only on $\deg P$, using the notation of that proposition (the error term corresponds to the contribution of those q such that $q\,(\mathrm{mod}\,p)$ is in one of the residue classes in the set S of Proposition 6.7; its size is bounded in terms of $\deg P$ only.)

Using the asymptotic formula (6.3) for the constant c_1, we get

$$\sum_{a\in P(\mathbf{F}_p)} E(X;p,a) = \sum_{2\leqslant i\leqslant k}\sum_{q\leqslant X} K_i(q) + O(p^{-1/2}X),$$

and Theorem 1.5 concludes the proof.

5.2. **Large Kloosterman sums with almost prime modulus.** In this section we prove Corollary 1.13. It is sufficient to prove the following:

Proposition 5.1. *For any $m\geqslant 2$, and δ such that $0<\delta<1/2$, there exists a constant $\beta_m>0$ such that*

$$|\{(p,q),\ p,q \textit{ primes } \geqslant X^{\delta},\ pq\leqslant X,\ |\mathrm{Kl}_m(1;pq)|\geqslant\beta_m\}| \gg \frac{X}{\log X}.$$

here the implicit constants depend on m and δ only.

We recall first the basic strategy from [33]. By the Chinese remainder theorem, we have the twisted multiplicativity

$$\mathrm{Kl}_m(1;pq) = \mathrm{Kl}_m(\overline{q}^m;p)\,\mathrm{Kl}_m(\overline{p}^m;q), \tag{5.1}$$

when p and q are distinct primes. Therefore, in order to prove the existence of pairs of primes (p,q) for which $|\,\mathrm{Kl}_m(1;pq)|$ is large, it is sufficient to show that there exists two sets of pairs of primes for which $|\,\mathrm{Kl}_m(\overline{q}^m;p)|$ and $|\,\mathrm{Kl}_m(\overline{p}^m;q)|$ are both large, and that these two sets intersect non-trivially. This leads us to proving that, for pairs (p,q) in suitable ranges, the hyper-Kloosterman sums $\mathrm{Kl}_m(\overline{q}^m;p)$ and $\mathrm{Kl}_m(\overline{p}^m;q)$ become equidistributed in the interval $[-m,m]$ with respect to a suitable measure. Such a statement is an instance of the vertical (or average) Sato-Tate laws of Katz and Deligne, but specialized to prime arguments.

To state properly these equidistribution statements, we recall that for any prime number p and auxiliary prime $\ell\neq p$, and for any isomorphism $\iota:\overline{\mathbf{Q}_\ell}\hookrightarrow\mathbf{C}$, there exists a $\overline{\mathbf{Q}}_\ell$-adic sheaf $\mathcal{K}\ell_m$ on $\mathbf{P}^1_{\mathbf{F}_p}$ (constructed by Deligne and studied by Katz in [29]) such that:

(1) The sheaf $\mathcal{K}\ell_m$ has rank m and is lisse on $\mathbf{G}_{m,\mathbf{F}_p}$, tamely ramified at 0 with a single Jordan block and wildly ramified at ∞ with Swan conductor 1 (in particular, we have $\mathrm{cond}(\mathcal{K}\ell)=m+3$);

(2) The sheaf $\mathcal{K}\ell_m$ is geometrically irreducible, and its geometric monodromy group is equal to $\mathrm{G}_m=\mathrm{SL}_m$ or Sp_m depending on whether m is odd or even;

(3) The sheaf $\mathcal{K}\ell_m$ is pointwise pure of weight 0, and for any $a\in\mathbf{F}_p^\times$, the trace of the Frobenius at a equals

$$\iota(\mathrm{tr}(\mathrm{Fr}_a\,|\mathcal{K}\ell_m)) = (-1)^{m-1}\,\mathrm{Kl}_m(a;p),$$

and moreover, for any choice of maximal compact subgroup K_m of $\mathrm{G}_m(\mathbf{C})$, $(\mathrm{Fr}_a\,|\mathcal{K}\ell_m)$ defines a unique conjugacy class $g^\natural_m(a;p)$ in K_m with trace equal to $(-1)^{m-1}\,\mathrm{Kl}_m(a;p)$.

It will be easy to prove the following result using Theorem 1.5:

Theorem 5.2 (Sato-Tate equidistribution). *Given $\delta<0$, $A\geqslant 1$, $P,Q\geqslant 2$ such that*

$$P^{3/4+\delta}\leqslant Q\leqslant P^A,$$

the set of conjugacy classes

$$\{g^\natural_m(\overline{q}^m;p),\ p\neq q \textit{ primes},\ (p,q)\in[P,2P]\times[Q,2Q]\}\subset K^\natural_m,$$

becomes equidistributed as $P\to+\infty$ with respect to the probability Haar measure μ_m on $K^\natural_m$.

Remark 5.3. A similar Sato-Tate equidistribution result over the primes holds for the generalized Kloosterman sheaves of Heinloth, Ngô and Yun [23] already mentioned in Remark 1.2.

We will sketch the proof below, but for the moment we can conclude from this the proof of Corollary 1.13. We pick $\alpha_m > 0$ small enough such that

$$\mu_m(\{g^\natural \in K_m^\natural,\ |\operatorname{tr}(g^\natural)| \geqslant \alpha_m\}) \geqslant 0.51,$$

(such an α_m exists because the direct image of the measure μ_m under the trace map $g^\natural \mapsto |\operatorname{tr}(g^\natural)|$ is absolutely continuous with respect to the Lebesgue probability measure on $[0, m]$).

Now let $\delta > 0$ be given, let P be large enough and consider Q such that

$$P^{3/4+\delta} \leqslant Q \leqslant P^{4/3-\delta}.$$

We then have

$$Q^{3/4+\delta'} \leqslant P \leqslant Q^{4/3-\delta'}$$

for some $\delta' > 0$ depending only on δ, and we can apply Theorem 5.2 twice to show that *both* sets

$$\mathcal{P}_1 = \{(p,q) \in [P,2P] \times [Q,2Q],\ p \neq q \text{ primes},\ |\operatorname{tr}(g_m^\natural(\overline{q}^m;p))| \geqslant \alpha_m\}$$
$$\mathcal{P}_2 = \{(p,q) \in [P,2P] \times [Q,2Q],\ p \neq q \text{ primes},\ |\operatorname{tr}(g_m^\natural(\overline{p}^m;q))| \geqslant \alpha_m\}$$

satisfy, as $P \to +\infty$, the limit

$$\frac{|\mathcal{P}_i|}{|\{(p,q) \in [P,2P] \times [Q,2Q]\}|} \longrightarrow \mu_m(\{g^\natural \in K_m^\natural,\ |\operatorname{tr}(g^\natural)| \geqslant \alpha_m\}) \geqslant 0.51.$$

In particular, the two sets have a non-empty intersection for P large enough, and in fact

$$|\mathcal{P}_1 \cap \mathcal{P}_2| \gg \frac{P}{\log P}\frac{Q}{\log Q}.$$

By (5.1), it follows that

$$|\{(p,q) \in [P,2P] \times [Q,2Q],\ p \neq q \text{ primes},\ |\operatorname{Kl}_m(1;pq)| \geqslant \alpha_m^2\}| \gg \frac{P}{\log P}\frac{Q}{\log Q}.$$

Then we obtain by an easy argument of dyadic partition that for X large enough, we have pairs

$$|\{(p,q),\ p \neq q \text{ primes, },\ p,q \geqslant X^{4/9},\ pq \in [X,2X],\ |\operatorname{Kl}_m(1;pq)| \geqslant \alpha_m^2\}| \gg \frac{X}{\log X},$$

as claimed.

Proof of Theorem 5.2. This is a direct application of the Weyl criterion. Let

$$X_{P,Q} = \{p \neq q \text{ primes},\ (p,q) \in [P,2P] \times [Q,2Q]\}.$$

It is enough to prove that if ϱ is a non-trivial irreducible representation of G_m, we have

$$\frac{1}{|X_{P,Q}|}\sum_{(p,q)\in X_{P,Q}} \operatorname{tr} \varrho(g_m^\natural(\overline{q}^m;p)) \longrightarrow 0 \tag{5.2}$$

as $P \to +\infty$.

Now, for each p, we can interpret the sum over q as the sum of the weight

$$K_\varrho(q) = \operatorname{tr} \varrho(g_m^\natural(\overline{q}^m;p))$$

modulo p. Now we claim that, for each $\varrho \neq 1$, the weight K_ϱ is a non-exceptional irreducible trace weight modulo p with conductor bounded by a constant depending only on m and ϱ. Assuming this, Theorem 1.5 (see 1.3) gives

$$\sum_{(p,q)\in X_{P,Q}} \operatorname{tr} \varrho(g_m^\natural(\overline{q}^m;p)) \ll \frac{PQ}{\log P}P^{-\eta}\Big(1 + \frac{P}{Q}\Big)^{1/12}$$

for any $\eta < 1/48$. Dividing by $|X_{P,Q}| \asymp PQ/(\log P)(\log Q)$, we get

$$\frac{1}{|X_{P,Q}|} \sum_{(p,q)\in X_{P,Q}} \operatorname{tr} \varrho(g_m^{\natural}(\overline{q}^m;p)) \ll (\log Q)(1+P/Q)^{1/2}P^{-\eta},$$

which tends to 0 provided $P^{3/4+\delta} < Q < P^A$ for some $\delta > 0$, $A \geqslant 1$.

To check the claim, we first define

$$\mathcal{K}\ell'_m = [x \mapsto x^{-m}]^* \mathcal{K}\ell$$

so that, for $a \in \mathbf{F}_p^{\times}$, we have the trace function

$$\iota((\operatorname{tr} \mathcal{K}\ell'_m)(\mathbf{F}_p, a)) = (-1)^{m-1} \operatorname{Kl}_m(a^{-m};p).$$

The function

$$K_{\varrho} : a \mapsto \operatorname{tr}(\varrho(g_m^{\natural}(a^{-m};p)))$$

is then (the restriction to $\mathbf{F}_p^{\times}$ of) the irreducible trace weight associated to the sheaf $\varrho(\mathcal{K}\ell')$ obtained by composing the representation $\mathcal{K}\ell'$ with the representation ϱ. In particular, this sheaf is also lisse and geometrically irreducible on $\mathbf{G}_m$, and of rank $\dim \varrho$. It is tame at ∞, and its Swan conductor at 0 is bounded in terms of m and $\dim \varrho$ only (by bounding the largest slope, see e.g. [33]), so the conductor is bounded in terms of m and $\deg \varrho$ only. Finally, because $\varrho(\mathcal{K}\ell')$ is irreducible of rank $\deg \varrho \geqslant 2$ (we use here the fact that both SL_m and Sp_m have no non-trivial representations of dimension 1), it follows that $\varrho(\mathcal{K}\ell')$ is not p-exceptional. □

6. Results from algebraic geometry

6.1. **Properties of the Fourier-Möbius group.** The goal of this section is to prove Proposition 3.1. In order to do so, we must first recall the definition of the Fourier-Möbius group of an isotypic sheaf $\mathcal{F}$, and establish a few of its properties which were not necessary in [12].

Let p be a prime. Let $\ell \neq p$ be an auxiliary prime number, $\iota : \bar{\mathbf{Q}}_\ell \simeq \mathbf{C}$ an isomorphism. Let ψ be the ℓ-adic additive character such that $\iota(\psi(x)) = e(x/p)$ for $x \in \mathbf{F}_p$.

Given any middle-extension sheaf $\mathcal{F}$ on $\mathbf{A}^1_{\mathbf{F}_p}$, any finite extension $k/\mathbf{F}_p$ and any $x \in \mathbf{P}^1(k)$, we denote by

$$(\operatorname{tr} \mathcal{F})(k,x)$$

the trace of the geometric Frobenius of k acting on the stalk of $\mathcal{F}$ at x. We also denote by $\mathrm{D}(\mathcal{F})$ the middle-extension dual of $\mathcal{F}$ given by $j_*(j^*\check{\mathcal{F}})$, where $j : U \hookrightarrow \mathbf{P}^1$ is the inclusion of any dense open set U on which $\mathcal{F}$ is lisse.

If $\mathcal{F}$ is any Fourier sheaf (in the sense of [29, Def. 8.2.2]) on $\mathbf{A}^1_{\mathbf{F}_p}$, we denote by $\mathrm{FT}(\mathcal{F})$ the Fourier transform of $\mathcal{F}$, computed by means of ψ, which satisfies

$$(\operatorname{tr} \mathrm{FT}(\mathcal{F}))(\mathbf{F}_p, y) = -\sum_{x\in\mathbf{F}_p} (\operatorname{tr} \mathcal{F})(\mathbf{F}_p, x)\psi(xy)$$

for any $y \in \mathbf{F}_p$. It is known that $\mathrm{FT}(\mathcal{F})$ is geometrically isotypic (resp. geometrically irreducible) if $\mathcal{F}$ is.

Let now $\mathcal{F}$ be an isotypic trace sheaf modulo p as in Definition 1.3. In [12], we defined the Fourier-Möbius group of $\mathcal{F}$ by

$$\mathbf{G}_{\mathcal{F}} = \{\gamma \in \mathrm{PGL}_2(\bar{\mathbf{F}}_p) \mid \gamma^*(\mathrm{FT}(\mathcal{F})) \simeq \mathrm{FT}(\mathcal{F})\},$$

where $\simeq$ denotes geometric isomorphism (see [12, Def. 1.14]). Furthermore, we defined the correlation sums of $\mathcal{F}$ by

$$\mathcal{C}(\mathcal{F};\gamma) = \frac{1}{p}\sum_{x\in\mathbf{F}_p} (\operatorname{tr} \mathrm{FT}(\mathcal{F}))(\mathbf{F}_p, \gamma\cdot x)\overline{(\operatorname{tr} \mathrm{FT}(\mathcal{F}))(\mathbf{F}_p, x)}$$

for $\gamma \in \mathrm{PGL}_2(\mathbf{F}_p)$.

The crucial link between these two notions is the following result (see [12, Cor. 9.2]) which follows from the Riemann Hypothesis over finite fields, and from bounds for the conductor of the Fourier transforms of Fourier sheaves.

Proposition 6.1. *Let p be a prime and let $\mathcal{F}$ be an isotypic trace sheaf modulo p. There exists $M \geqslant 1$, depending only polynomially on* $\mathrm{cond}(\mathcal{F})$*, such that*

$$|\iota(\mathcal{C}(\mathcal{F};\gamma))| \leqslant M\sqrt{p}$$

for all $\gamma \notin \mathbf{G}_{\mathcal{F}}$.

Let then

$$\mathbf{B}_{\mathcal{F}} = \mathbf{G}_{\mathcal{F}} \cap \mathbf{B},$$

where $\mathbf{B} \subset \mathrm{PGL}_2$ is the upper-triangular Borel subgroup. We deduce from the proposition above:

Proposition 6.2. *Let p be a prime, let $\mathcal{F}$ be an isotypic trace sheaf modulo p, and let $K(x) = \iota((\mathrm{tr}\,\mathcal{F})(\mathbf{F}_p, x))$ denote the trace function of $\mathcal{F}$ on $\mathbf{F}_p$. There exists $M \geqslant 1$, depending only polynomially on* $\mathrm{cond}(\mathcal{F})$*, such that*

$$\Bigl|\sum_{x\in\mathbf{F}_p} K(x)\overline{K(ax)}e\Bigl(\frac{bx}{p}\Bigr)\Bigr| \leqslant M\sqrt{p}$$

if

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \notin \mathbf{B}_{\mathcal{F}}. \tag{6.1}$$

Proof. By means of the Plancherel formula for the finite-field Fourier transform, we check easily that

$$\sum_{x\in\mathbf{F}_p} K(x)\overline{K(ax)}e\Bigl(\frac{bx}{p}\Bigr) = \iota(\mathcal{C}(\mathcal{F};\gamma))$$

where γ is the upper-triangular matrix in (6.1). Hence the proposition gives the result. □

It follows now that Proposition 3.1 is a consequence of the next theorem:

Theorem 6.3. *Let p be a prime and let $\mathcal{F}$ be an isotypic sheaf. At least one of the following four properties holds:*

(1) *The trace function of $\mathcal{F}$ is proportional to the trace function of a sheaf $\mathcal{L}_{\psi(aX)}$ for some $a \in \mathbf{F}_p$, i.e., to an additive character, or to a delta function at some point $a \in \mathbf{F}_p$;*

(2) *The group $\mathbf{B}_{\mathcal{F}}$ has dimension $\geqslant 1$ and $\mathcal{F}$ is p-exceptional, i.e., its unique geometrically irreducible component is a tensor product $\mathcal{L}_\chi \otimes \mathcal{L}_\eta$ for some non-trivial Kummer sheaf $\mathcal{L}_\chi$ and some possibly trivial additive character η;*

(3) *The group $\mathbf{B}_{\mathcal{F}}$ is finite and*

$$|\mathbf{B}_{\mathcal{F}}(\mathbf{F}_p)| \leqslant 10\,\mathrm{cond}(\mathcal{F})^2;$$

(4) *The conductor of $\mathcal{F}$ is at least $(p/10)^{1/2}$.*

To prove this, we first prove two basic properties of the Fourier-Möbius group and one lemma concerning Swan conductors.

Proposition 6.4. *Let k be a finite field, and let $\mathcal{F}$ be an ℓ-adic isotypic trace sheaf on $\mathbf{A}^1_k$. Let $\mathcal{G}$ be its Fourier transform. Then the subgroup $\mathbf{G}_{\mathcal{F}} \subset \mathrm{PGL}_2(\bar{k})$ is an algebraic subgroup defined over k.*

In particular, for $\mathcal{F}$ over $\mathbf{F}_p$, $\mathbf{B}_{\mathcal{F}}$ is an algebraic subgroup of $\mathbf{B}$ defined over $\mathbf{F}_p$.

We thank R. Pink for explaining to us how to prove this proposition.

Proof. Let $S \subset \mathbf{P}^1$ be the divisor of singularities of $\mathcal{G}$, so that $U = \mathbf{P}^1 - S$ is the largest open set on which it is lisse. Because $\mathcal{G}$ is non-constant (the sheaf $\mathcal{F}$ would have to be a Dirac delta sheaf supported on a single point for this to happen, and such a sheaf is not a Fourier sheaf), we have $S \neq \emptyset$. Let $G \subset \mathrm{PGL}_2$ be the stabilizer of S, which is a proper algebraic subgroup of PGL_2 defined over $\mathbf{F}_p$. Then we have a first inclusion $\mathbf{G}_{\mathcal{F}} \subset G$.

Now we work over $\bar{k}$, and just denote by U its base-change to $\bar{k}$. We consider the action morphism

$$\mu : \begin{cases} G \times U \longrightarrow U \\ (\gamma, x) \mapsto \gamma \cdot x \end{cases}$$

and the second projection $p_2 : G \times U \longrightarrow U$, and we define the sheaf

$$\mathcal{E} = \mu^* \mathcal{G} \otimes p_2^* \,\mathrm{D}(\mathcal{G})$$

on $G \times U$ and the higher direct-image $\mathcal{I} = R^2 p_{1,!}\mathcal{E}$, which is a sheaf on the algebraic group $G/\bar{k}$. By the base-change theorem for higher-direct images with compact support [7, Arcata, IV, Th. 5.4], the stalk of $\mathcal{I}$ at a geometric point $\gamma \in G(\bar{k})$ is naturally isomorphic to $H^2_c(U, \gamma^*\mathcal{G} \otimes \mathrm{D}(\mathcal{G}))$.

Furthermore, the constructibility theorem for higher direct images with compact support [7, Arcata, IV, Th. 6.2] shows that $\mathcal{I}$ is a constructible ℓ-adic sheaf on G. This implies (see also [7, Rapport, Prop. 2.5]) that for any $d \geqslant 0$, the set

$$\{\gamma \in G(\bar{k}) \mid \dim \mathcal{I}_\gamma = \dim H^2_c(U, \gamma^*\mathcal{G} \otimes \mathrm{D}(\mathcal{G})) = d\}$$

is constructible in $G(\bar{k})$, i.e., is a finite union of locally-closed subsets. In particular, the set of all γ where

$$H^2_c(U, \gamma^*\mathcal{G} \otimes \mathrm{D}(\mathcal{G})) \neq 0,$$

is constructible. But this set is exactly $\mathbf{G}_{\mathcal{F}}$ by the co-invariant formula for H^2_c on a curve (see [12, Th. 9.1]). Since it is well-known that a constructible subgroup of an algebraic group is Zariski-closed (see, e.g., [2, Ch. I, Prop. 1.3]) we conclude therefore that $\mathbf{G}_{\mathcal{F}}$ is a closed subgroup of PGL_2. □

Next we need to understand when $\mathbf{G}_{\mathcal{F}}$ can be "large". We prove here a bit more than what we need for the sake of completeness. We use the notation $\mathrm{T}^{x,y}$ for the maximal torus in PGL_2 defined as the pointwise stabilizer of $\{x, y\} \subset \mathbf{P}^1$ (for $x \neq y$) and U^x for the unipotent radical of the Borel subgroup $\mathbf{B}^x$ which is the stabilizer of $x \in \mathbf{P}^1$.

Proposition 6.5. *Let $\mathcal{F}$ be a geometrically isotypic ℓ-adic Fourier sheaf on $\mathbf{A}^1_{\mathbf{F}_p}$, with Fourier transform $\mathcal{G} = \mathrm{FT}_\psi(\mathcal{F})$ with respect to some non-trivial additive character ψ.*

(1) *If there exists $x \in \mathbf{P}^1$ such that $\mathbf{G}_{\mathcal{F}} \supset \mathrm{U}^x$, then $\mathcal{G}$ is geometrically isomorphic to a direct sum of copies of $\mathcal{L}_{\psi_0(\gamma_0(X))}$ for some non-trivial additive character ψ_0, where $\gamma_0 \in \mathrm{PGL}_2$ is such that $\gamma_0 \cdot x = \infty$. In that case, we have $\mathbf{G}_{\mathcal{F}} = \mathrm{U}^x$.*

(2) *If there exist $x \neq y$ in $\mathbf{P}^1$ such that $\mathbf{G}_{\mathcal{F}} \supset \mathrm{T}^{x,y}$, then $\mathcal{G}$ is geometrically isomorphic to a direct sum of copies of $\mathcal{L}_{\chi_0(\gamma_0(X))}$ for some non-trivial multiplicative character χ_0, where $\gamma_0 \in \mathrm{PGL}_2$ is such that $\gamma_0 \cdot x = 0$, $\gamma_0 \cdot y = \infty$. In that case, we have $\mathbf{G}_{\mathcal{F}} = \mathrm{T}^{x,y}$ if χ_0 is not of order 2, and $\mathbf{G}_{\mathcal{F}} = \mathrm{N}^{x,y}$, the normalizer of $\mathrm{T}^{x,y}$, if $\chi_0^2 = 1$.*

Proof. (1) The "if" direction is immediate. For the converse, we may first assume that $x = \infty$, by conjugation with a matrix γ_0 with $\gamma_0 \cdot x = \infty$. The assumption is then that

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}^* \mathcal{G} \simeq \mathcal{G},$$

for any $t \in \bar{\mathbf{F}}_p$, where the symbol $\simeq$ denotes geometric isomorphism. Since $\mathcal{G}$ is geometrically isotypic, we also have

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}^* \mathcal{G}_1 \simeq \mathcal{G}_1,$$

for $t \in \bar{\mathbf{F}}_p$, where $\mathcal{G}_1$ is the geometrically irreducible component of $\mathcal{G}$. We can then apply [30, Lemma 2.6.13] to deduce that

$$\mathcal{G}_1 \simeq \mathcal{L}_{\psi_0(X)}$$

(geometrically) for some additive ℓ-adic character ψ_0, and hence $\mathcal{G}$ is a direct sum of copies of this Artin-Schreier sheaf. Furthermore, it follows from the classification of Artin-Schreier sheaves that if ψ_0 is non-trivial and $\gamma \notin \mathrm{U}^\infty$, we do not have $\gamma^* \mathcal{L}_{\psi_0(X)} \simeq \mathcal{L}_{\psi_0(X)}$, we hence the Fourier-Möbius group is exactly equal to U^∞.

(2) As before, we may first conjugate using some γ_0 to reduce to the case where $x = 0$, $y = \infty$, and we may reduce to the case where $\mathcal{F}$ and $\mathcal{G}$ are geometrically irreducible, so that the assumption is

$$\mathbf{G}_{\mathcal{F}} \supset \mathrm{T} = \mathrm{T}^{0,\infty} = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \right\}$$

for all a, $d \in \bar{k}$. By [30, Lemma 2.6.13], again, there exists a multiplicative character χ_0 such that

$$\mathcal{G} \simeq \mathcal{L}_{\chi_0(X)}.$$

This character is non-trivial since $\mathcal{G}$ is a Fourier sheaf. Now to finish the computation of $\mathbf{G}_{\mathcal{F}}$, we use the fact that $\mathcal{L}_{\chi_0(X)}$ is tamely ramified at 0 and ∞, and hence

$$\mathbf{G}_{\mathcal{F}} \subset \mathrm{N} = \mathrm{N}^{0,\infty} = \mathrm{T} \cup \left\{ \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix} \right\},$$

the normalizer of T in PGL_2. Clearly, $\mathrm{T} \subset \mathbf{G}_{\mathcal{L}_{\chi_0(X)}}$. If $\gamma \in \mathrm{N} - \mathrm{T}$, on the other hand, we have $\gamma^* \mathcal{F}_{\chi_0(X)} \simeq \mathcal{F}_{\chi_0(X^{-1})}$, and by the classification of Kummer sheaves, it follows that $\gamma \in \mathbf{G}_{\mathcal{L}_{\chi_0(X)}}$ if and only if $\chi_0 = \chi_0^{-1}$, i.e., if χ_0 is of order 2. □

The second lemma concerns the size of Swan conductors of lisse sheaves on $\mathbf{G}_m$ with some non-trivial (multiplicative) translation-invariance property.

Lemma 6.6. *Let k be an algebraic closure of a finite field of characteristic p, and let $\mathcal{F}$ be an ℓ-adic sheaf for some $\ell \neq p$ which is lisse on $\mathbf{G}_{m,k}$. If there exists $a \neq 1$ in $\mathbf{G}_m(k)$ such that $\mathcal{F} \simeq [\times a]^* \mathcal{F}$, then $m \mid \mathrm{Swan}_\infty(\mathcal{F})$, where m is the multiplicative order of a. In particular, if $\mathcal{F}$ is not tame at ∞, we have* $\mathrm{Swan}_\infty(\mathcal{F}) \geqslant m$.

Proof. Let V be the generic stalk of $\mathcal{F}$, seen as a representation of the inertia group $I = I(\infty)$ at ∞, and let

$$V = \bigoplus_{\alpha \in A} V_\alpha$$

be the decomposition of V in I-isotypic subspaces. Let W_α denote the irreducible I-representation such that V_α is a multiple of W_α.

The finite cyclic subgroup $G \subset \mathbf{G}_m(k)$ of order m generated by a acts on the index set A, corresponding to the fact that $[\times a]^* V = V$ as I-representation: we have

$$[\times a^j]^* V_\alpha = V_{a^j \cdot \alpha},$$

for any integer $j \geqslant 0$, and in fact even

$$[\times a^j]^* W_\alpha = W_{a^j \cdot \alpha},$$

since W_α is uniquely determined by V_α.

Let $B \subset A$ be one of the orbits of G. Its size $|B|$ is a divisor of m, and if $\alpha \in B$, we have an isomorphism $[\times a^{|B|}]^* W_\alpha = W_\alpha$. Since W_α is irreducible, we can apply [29, Prop. 4.1.6 (2)] to deduce that

$$\mathrm{Swan}_\infty(W_\alpha) \equiv 0 \,(\mathrm{mod}\, m/|B|).$$

Since multiplicative translation by a is an automorphism, it follows that

$$\mathrm{Swan}_\infty(W_{a^j \cdot \alpha}) = \mathrm{Swan}_\infty(W_\alpha) \equiv 0 \,(\mathrm{mod}\, m/|B|)$$

for any $a^j \in G$. Summing over the orbit, we get

$$\mathrm{Swan}_\infty\Big(\bigoplus_{\alpha \in B} V_\alpha\Big) \equiv 0 \,(\mathrm{mod}\, m),$$

and then summing over the orbits we get

$$\mathrm{Swan}_\infty(V) \equiv 0 \,(\mathrm{mod}\, m).$$

If $\mathcal{F}$ is wild at infinity, than $\mathrm{Swan}_\infty(V) \neq 0$, and therefore it must be $\geqslant m$. □

Having dealt with these preliminaries, we can now prove the theorem.

Proof of Theorem 6.3. The group $B = \mathbf{B}_{\mathcal{F}}(\mathbf{F}_p)$ is a finite subgroup of $\mathbf{B} \cap \mathrm{PGL}_2(\mathbf{F}_p)$. We distinguish three situations in turn.

(1) If B contains a non-trivial unipotent element g, then since g fixes ∞, the reasoning in [12, §9, Proof of Th. 1.12] shows that either $\mathrm{cond}(\mathcal{G}) \geqslant p$, in which case the fourth case holds by [12, Prop. 8.2 (1)], or otherwise the trace function of the Fourier transform $\mathrm{FT}_\psi(\mathcal{F})$ is proportional to an additive character, so that the trace function of $\mathcal{F}$ is proportional to a delta function, and we are in the first case.

Now, if B contains no unipotent elements, the unipotent radical of $\mathbf{B}_{\mathcal{F}}$ must also be trivial (otherwise it would have non-trivial $\mathbf{F}_p$-points). So, by the structure of $\mathbf{B}$, the connected component of $\mathbf{B}_{\mathcal{F}}$ is contained in a conjugate (say $\mathbf{D}$) of the diagonal subgroup in $\mathbf{B}$. We continue with two further possibilities:

(2) If $\mathbf{B}_{\mathcal{F}} \supset \mathbf{D}$, we deduce from Proposition 6.5 (2) that the Fourier transform of $\mathcal{F}$ is geometrically isomorphic to a direct sum of copies of $\mathcal{L}_{\chi(\gamma(X))}$ for some multiplicative character χ and some $\gamma \in \mathbf{B}$. By Fourier transform, this implies that $\mathcal{F}$ is geometrically isomorphic to a direct sum of copies of the tensor product $\mathcal{L}_\chi \otimes \mathcal{L}_\eta$ for some multiplicative character χ and some additive character η. Here χ must be non-trivial because otherwise $\mathcal{F}$ would not be a Fourier sheaf, and we are in the second case of the statement of the proposition.

(3) Finally, if $\mathbf{B}_{\mathcal{F}}$ is a finite group, its finite subgroup $B \subset \mathbf{D}$ is cyclic, and there exists $x_0 \in \mathbf{A}^1$ such that all elements of B fix ∞ and x_0. Let $\mathcal{G}$ be the Fourier transform of $\mathcal{F}$. Replacing $\mathcal{G}$ with $\mathcal{G}_0 = [-x_0]^*\mathcal{G}$, which has the same conductor as $\mathcal{G}$, we can assume that $x_0 = 0$, and hence that B can be identified with a finite cyclic subgroup of $\mathbf{F}_p^\times$ acting on $\mathbf{P}^1$ by multiplication. Let $a \in \mathbf{F}_p^\times$ be a generator of $B \subset \mathbf{F}_p^\times$. There are two subcases:

– (3.1) If $\mathcal{G}_0$ is not lisse on $\mathbf{G}_m$, there is a non-zero singularity $s \in \mathbf{G}_m$ of $\mathcal{G}_0$; the geometric isomorphism $\mathcal{G}_0 \simeq [\times a]^*\mathcal{G}_0$ implies that the orbit of s under multiplication by powers of a is also contained in the set S of singularities of $\mathcal{G}_0$. This set contains $\geqslant |B|$ elements, and therefore

$$\mathrm{cond}(\mathcal{G}) = \mathrm{cond}(\mathcal{G}_0) \geqslant |S| \geqslant |B|$$

in that case, and by [12, Prop. 8.2 (1)], we get

$$|B| \leqslant \mathrm{cond}(\mathcal{G}) \leqslant 10\,\mathrm{cond}(\mathcal{F})^2,$$

i.e., case (3) of the theorem.

– (3.2) If $\mathcal{G}_0$ is lisse on $\mathbf{G}_m$, we first note that $\mathcal{G}_0$ can not be tame at both 0 and ∞, since the tame fundamental group of $\mathbf{G}_m$ is abelian and $\mathcal{G}_0$ would then be a Kummer sheaf, which we excluded by assuming that $\mathbf{B}_{\mathcal{F}}$ is finite (again from Proposition 6.5, (2)). Up to applying a further automorphism $x \mapsto x^{-1}$, we can assume that $\mathcal{G}_0$ is wildly ramified at ∞. We can then apply Lemma 6.6 to $\mathcal{G}_0$, and deduce that

$$\mathrm{Swan}_{\infty}(\mathcal{G}_0) \geqslant |B|,$$

and hence we get again

$$\mathrm{cond}(\mathcal{G}) = \mathrm{cond}(\mathcal{G}_0) \geqslant \mathrm{Swan}_{\infty}(\mathcal{G}_0) \geqslant |B|,$$

and conclude as before. □

6.2. **Decomposition of characteristic functions.** In this section, we explain the necessary properties of the trace weights underlying Corollary 1.10. We recall especially the decomposition of the characteristic function of the set of values $P(n)$ of a polynomial $P \in \mathbf{F}_p[X]$ with $n \in \mathbf{F}_p$ in terms of trace functions. These types of results are well-known, but we give the full proof since we require some quantitative information concerning this decomposition.

Proposition 6.7. *Let p be prime and let $P \in \mathbf{F}_p[X]$ be a non-constant polynomial of degree $\deg P < p$. Let $\mathcal{P}$ be the set of values of P modulo p and $\mathbf{1}_P$ be its characteristic function.*

There exist a finite set $S \subset \mathbf{F}_p$ with order at most $\deg P$, an integer $k \geqslant 1$ and a finite number of trace functions K_i associated to middle-extension sheaves $\mathcal{F}_i$, $1 \leqslant i \leqslant k$, which are pointwise pure of weight 0, and algebraic numbers $c_i \in \bar{\mathbf{Q}}$, such that

$$\sum_i c_i K_i(x) = \mathbf{1}_P(x) \tag{6.2}$$

for all $x \in \mathbf{F}_p - S$, and with the following properties:

- *The constants k, $|c_i|$ and $\mathrm{cond}(\mathcal{F}_i)$ are bounded in terms of $\deg P$ only;*
- *The sheaf $\mathcal{F}_1$ is trivial and none of the $\mathcal{F}_i$ for $i \neq 1$ are geometrically trivial, and furthermore*

$$c_1 = \frac{|\mathcal{P}|}{p} + O(p^{-1/2}), \tag{6.3}$$

where the implicit constant depends only on $\deg P$;

- *If P is squarefree, no $\mathcal{F}_i$, $i \neq 1$, contains an exceptional sheaf as a Jordan-Hölder factor.*

Proof. Let $K(x)$, for $x \in \mathbf{F}_p$, denote the characteristic function of the set of values $P(y)$ for $y \in \mathbf{F}_p$, so that we are trying to express K as a linear combination of trace weights.

Let $\tilde{D} \subset \mathbf{A}^1$ be the critical points of P, $\tilde{S} = P(\tilde{D}) \subset \mathbf{A}^1$ the set of critical values, so that P restricts to a finite étale covering

$$V = \mathbf{A}^1 - \tilde{D} \longrightarrow U = \mathbf{A}^1 - \tilde{S}$$

and let

$$W \xrightarrow{\pi} V \longrightarrow U$$

be the Galois closure of V. The Galois group $G = \mathrm{Gal}(W/U)$ contains the subgroup $H = \mathrm{Gal}(W/V)$, and has order dividing $\deg(P)!$, hence coprime to p.

For any $x \in U(\mathbf{F}_p)$, the Galois group G permutes the points of the fiber $\pi^{-1}(x) \subset W$, and this Galois action is isomorphic to the left-translation action on G/H. The Frobenius $\mathrm{Fr}_{x,p}$ at x, seen as an element of G, also permutes the points of the fiber, and the subset of rational points $\pi^{-1}(x) \cap W(\mathbf{F}_p)$ correspond bijectively to the fixed points of $\mathrm{Fr}_{x,p}$, and hence to the number of fixed points of $\mathrm{Fr}_{x,p}$ acting on G/H, which is equal to the number of conjugates of $\mathrm{Fr}_{x,p}$ that are in H.

More generally, if we consider the function

$$\theta\,:\,\begin{cases} G \longrightarrow \bar{\mathbf{Q}}_\ell \\ g \mapsto \begin{cases} 1 & \text{if } g \text{ is conjugate to } \textit{some } h \in H \\ 0 & \text{otherwise,} \end{cases} \end{cases}$$

the same argument implies that we have

$$K(x) = \theta(\mathrm{Fr}_{x,p})$$

for all $x \in U(\mathbf{F}_p)$.

The function θ is invariant under G-conjugation. Hence, by character theory (since $\ell \neq p$, the $\bar{\mathbf{Q}}_\ell$-linear representations of G can be identified with the $\mathbf{C}$-linear representations) there exist coefficients c_ϱ such that

$$\theta = \sum_\varrho c_\varrho \chi_\varrho$$

where ϱ runs over isomorphism classes of irreducible $\bar{\mathbf{Q}}_\ell$-linear representations

$$\varrho\,:\,G \longrightarrow \mathrm{GL}(V_\varrho)$$

of G and $\chi_\varrho = \mathrm{tr}\,\varrho$ denotes the character of ϱ. By composition

$$\Lambda_\varrho\,:\,\pi_1(U) \longrightarrow \pi_1(U)/\pi_1(W) \simeq G \xrightarrow{\varrho} \mathrm{GL}(V_\varrho)$$

each ϱ determines an ℓ-adic lisse sheaf Λ_ϱ on U which is pointwise pure of weight 0 and satisfies

$$\chi_\varrho(\mathrm{Fr}_{x,p}) = (\mathrm{tr}\,\Lambda_\varrho)(x, \mathbf{F}_p)$$

for all $x \in U(\mathbf{F}_p)$. We therefore obtain

$$K(x) = \sum_\varrho c_\varrho K_\varrho$$

for $x \in U(\mathbf{F}_p)$, where K_ϱ is the trace function of Λ_ϱ.

We rearrange this slightly for convenience. Let $\mathcal{T}$ denote the set of ϱ such that Λ_ϱ is geometrically trivial. We know that K_ϱ is a constant of weight 0, say α_ϱ, for $\varrho \in \mathcal{T}$, and we define $\mathcal{F}_1 = \bar{\mathbf{Q}}_\ell$, so $K_1(x) = 1$, and

$$c_1 = \sum_{\varrho \in \mathcal{T}} c_\varrho \alpha_\varrho.$$

Then we enumerate arbitrarily

$$\{\varrho \notin \mathcal{T}\} = \{\varrho_2, \ldots, \varrho_k\}$$

and take $\mathcal{F}_i = \Lambda_{\varrho_i}$, $c_i = c_{\varrho_i}$. This gives the desired decomposition (6.2) with $S = \tilde{S}(\mathbf{F}_p)$, which has $\leqslant |\tilde{S}| \leqslant \deg P$ elements.

We now bound the numerical invariants in this decomposition. First, note that the number of non-zero summands is at most the number of ϱ, i.e, the number of conjugacy classes in G, and hence is bounded in terms of $\deg P$ only. For any ϱ we have

$$|c_\varrho| = \Big|\frac{1}{|G|}\sum_{g\in G} \theta(g)\chi_\varrho(g)\Big| \leqslant \dim \varrho \leqslant \sqrt{|G|}$$

which is bounded in terms of $\deg P$ only (using very trivial bounds $|\chi_\varrho(g)| \leqslant \dim \varrho$, $|\theta(g)| \leqslant 1$ and the fact that the sum of squares of $\dim \varrho$ is equal to $|G|$). And since $p \nmid |G|$, all sheaves Λ_ϱ are tame, and since they are unramified outside S, we get

$$\mathrm{cond}(\Lambda_\varrho) \leqslant |S| + \dim \varrho$$

which is again bounded in terms of $\deg P$ only.

Moreover, none of the sheaves Λ_ϱ can contain a Jordan-Hölder factor geometrically isomorphic to $\mathcal{L}_{\chi(X)} \otimes \mathcal{L}_{\psi(X)}$ with ψ non-trivial, since the Λ_ϱ are tamely ramified everywhere. If we assume that P is squarefree, 0 is not a critical value, and all the sheaves Λ_ϱ are unramified at 0 and therefore cannot have a non-trivial Kummer sheaf as (geometric) Jordan-Hölder factor. Thus the sheaf $\mathcal{F}_i$ does not contain an exceptional factor in this case.

We conclude by proving (6.3): we have

$$\begin{aligned}|\mathcal{P} \cap U(\mathbf{F}_p)| &= \sum_{x\in U(\mathbf{F}_p)} \theta(\mathrm{Fr}_{x,p}) \\ &= \sum_{\varrho} c_\varrho \sum_{x\in U(\mathbf{F}_p)} K_\varrho(x) \\ &= c_1|U(\mathbf{F}_p)| + \sum_{\varrho\notin\mathcal{T}} c_\varrho \sum_{x\in U(\mathbf{F}_p)} K_\varrho(x).\end{aligned}$$

For each ϱ which is not geometrically trivial, we can apply the Riemann Hypothesis to the inner sum, which shows it is $\ll p^{1/2}$ with an implicit constant that depends only on $\deg P$ (since the conductor of Λ_ϱ is bounded in terms of $\deg P$ only). Since the number of ϱ and the constants c_ϱ are also bounded in terms of $\deg P$ only, we obtain

$$c_1 = \frac{|\mathcal{P} \cap U(\mathbf{F}_p)|}{|U(\mathbf{F}_p)|} + O(p^{1/2}|U(\mathbf{F}_p)|^{-1}),$$

hence the result since $p - \deg P \leqslant |U(\mathbf{F}_p)| \leqslant p$. □

References

[1] V. Blomer, *Subconvexity for twisted L-functions on GL(3)*, Am. Journ. Math. **134** (2012), no. 5, 1385–1421.

[2] A. Borel, *Linear algebraic groups*, Graduate Texts in Math., vol. 126, Springer, 1991.

[3] J. Bourgain, *More on the sum-product phenomenon in prime fields and its applications*, Int. J. Number Theory **1** (2005), no. 1, 1–32.

[4] ———, *On the Fourier-Walsh spectrum of the Moebius function*, Israel J. of Math., to appear.

[5] J. Bourgain and M. Z. Garaev, *Sumsets of reciprocals in prime fields and multilinear Kloosterman sums*, `arXiv:1211.4184` (2012).

[6] J. Bourgain, P. Sarnak, and T. Ziegler, *Disjointness of Mobius from horocycle flows*, `arXiv:1110.0992`, Preprint.

[7] P. Deligne, *Cohomologie étale, SGA* 4 1/2, Lecture Notes in Mathematics, vol. 569, Springer, 1977.

[8] ———, *letter to V. Drinfeld*, dated June 18, 2011.

[9] H. Esnault and M. Kerz, *A finiteness theorem for Galois representations of function fields over finite fields (after Deligne)*, preprint `arXiv:1208.0128v1` (2012).

[10] É. Fouvry, *Autour du théorème de Bombieri-Vinogradov*, Acta Math. **152** (1984), no. 3-4, 219–244.

[11] ———, *Sur le problème des diviseurs de Titchmarsh*, J. reine angew. Math. **357** (1985), 51–76.

[12] É. Fouvry, E. Kowalski, and Ph. Michel, *Algebraic twists of modular forms and Hecke orbits*, Preprint `arXiv:1207.0617` (2012).

[13] ———, *Counting sheaves using spherical codes*, Preprint `arXiv:1210.0851` (2012).

[14] É. Fouvry and Ph. Michel, *Sur certaines sommes d'exponentielles sur les nombres premiers*, Ann. Sci. École Norm. Sup. (4) **31** (1998), no. 1, 93–130.

[15] É Fouvry and Ph. Michel, *Sur le changement de signe des sommes de Kloosterman*, Ann. of Math. (2) **165** (2007), no. 3, 675–715.

[16] É. Fouvry and Ph. Michel, *Sommes de modules de sommes d'exponentielles*, Pacific J. Math. **209** (2003), no. 2, 261–288, DOI 10.2140/pjm.2003.209.261 (French, with English summary).

[17] É. Fouvry and I. E. Shparlinski, *On a ternary quadratic form over primes*, Acta Arith. **150** (2011), no. 3, 285–314.

[18] J. B. Friedlander, K. Gong, and I. Shparlinskiĭ, *Character sums over shifted primes*, Mat. Zametki **88** (2010), no. 4, 605–619 (Russian, with Russian summary); English transl., Math. Notes **88** (2010), no. 3-4, 585–598.

[19] J. B. Friedlander and H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem*, Ann. of Math. (2) **121** (1985), no. 2, 319–350. With an appendix by B.J. Birch and E. Bombieri.

[20] B.J. Green, *On (not) computing the Moebius function using bounded depth circuits*, Combinatorics, Probability and Computing (to appear).

[21] G. Harman, *Trigonometric sums over primes. I*, Mathematika **28** (1981), no. 2, 249–254 (1982).

[22] D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), no. 6, 1365–1377.

[23] J. Heinloth, B.-C. Ngô, and Z. Yun, *Kloosterman sheaves for reductive groups*, Ann. of Math. (2012), to appear, available at `arXiv:1005.2765`.

[24] L. K. Hua, *Additive theory of prime numbers*, Translations of Mathematical Monographs, Vol. 13, American Mathematical Society, Providence, R.I., 1965.

[25] H. Iwaniec, *Introduction to the spectral theory of automorphic forms*, Biblioteca de la Revista Matemática Iberoamericana, Revista Matemática Iberoamericana, Madrid, 1995.

[26] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004.

[27] H. Iwaniec, W. Luo, and P. Sarnak, *Low lying zeros of families of L-functions*, Inst. Hautes Études Sci. Publ. Math. **91** (2000), 55–131 (2001).

[28] A. A. Karatsuba, *Sums of characters with prime numbers*, Izv. Akad. Nauk SSSR Ser. Mat. **34** (1970), 299–321.

[29] N. M. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Mathematics Studies, vol. 116, Princeton University Press, Princeton, NJ, 1988.

[30] ______, *Rigid local systems*, Annals of Mathematics Studies, vol. 139, Princeton University Press, Princeton, NJ, 1993.

[31] K. Matomäki, *A note on signs of Kloosterman sums*, Bull. Soc. Math. France **139** (2011), no. 3, 287–295.

[32] Ph. Michel, *Autour des conjectures de Sato-Tate*, Thèse de Doctorat ès Sciences, Université de Paris-Sud (1995).

[33] ______, *Autour de la conjecture de Sato-Tate pour les sommes de Kloosterman. I*, Invent. math. **121** (1995), no. 1, 61–78.

[34] ______, *Minorations de sommes d'exponentielles*, Duke Math. J. **95** (1998), no. 2, 227–240.

[35] H. L. Montgomery, *Topics in multiplicative number theory*, Lecture Notes in Mathematics, vol. 227, Springer, 1971.

[36] N. Pitt, *On an analogue of Titchmarsh's divisor problem for holomorphic cusp forms*, J. Amer. Math. Soc., posted on 2012, DOI http://dx.doi.org/10.1090/S0894-0347-2012-00750-4, (to appear in print).

[37] P. Sarnak, *Moebius randomness and dynamics* (2010), available at `http://publications.ias.edu/sarnak/section/514`.

[38] J. Sivak-Fischler, *Crible étrange et sommes de Kloosterman*, Acta Arith. **128** (2007), no. 1, 69–100, DOI 10.4064/aa128-1-4 (French).

[39] ______, *Crible asymptotique et sommes de Kloosterman*, Bull. Soc. Math. France **137** (2009), no. 1, 1–62.

[40] Z. Yun, *Examples of Kloosterman sheaves*, manuscript (2009).

Université Paris Sud, Laboratoire de Mathématique, Campus d'Orsay, 91405 Orsay Cedex, France
E-mail address: `etienne.fouvry@math.u-psud.fr`

ETH Zürich – D-MATH, Rämistrasse 101, CH-8092 Zürich, Switzerland
E-mail address: `kowalski@math.ethz.ch`

EPFL/SB/IMB/TAN, Station 8, CH-1015 Lausanne, Switzerland
E-mail address: `philippe.michel@epfl.ch`